Title of Pre-application: Essential computing and software development for the DUNE experiment.
Lead Principal Investigator: Heidi Schellman, Professor of Physics
Lead Institution: Oregon State University
Lead Principal Investigator Phone: 541 737 4631
and **email**: Heidi.Schellman@oregonstate.edu
DOE/SC Program Office: High Energy Physics
Funding Opportunity Announcement number: FOA Number: DE-FOA-0002181

Edited to remove unfunded projects, Dec. 13, 2021

The DUNE neutrino oscillation experiment will run in the middle of the next decade with commissioning of the data acquisition systems for the first far detector module expected to start in 2023-2024. ProtoDUNE prototype runs are ongoing and will continue through beam tests in 2021-22 at CERN. This proposal is for a 3-year effort to build and test a robust foundation for US DUNE computing prior to the start of commissioning of the first far detector module.

DUNE requires a global software and computing effort to store, catalog, reconstruct, calibrate and analyze approximately 30 PB of data/year from liquid argon detectors larger in scale than any previous neutrino experiment. Single event sizes are expected to range from 200 MB for protoDUNE, to 6 GB for a far detector module, to 460 TB for a full readout of a 100 s supernova candidate. Full sensitivity to neutrino oscillations and supernovae neutrinos require precise energy calibration and energy thresholds in the few MeV range, which present significant challenges for signal processing. In addition to the general distributed computing issues that are common to large HEP experiments, the very large event sizes present DUNE with a unique computing challenge. DUNE intends to benefit from previous experience and will contribute to ongoing improvements in general HEP computing infrastructure, working in collaboration with the OSG, the WLCG, and the HEP Software Foundation among others. However, the unique nature of DUNE events will require dedicated effort to adapt and integrate DUNE-specific solutions to achieve the physics goals of the experiment.

The purpose of this proposal is to build a cooperative US effort between US Universities and DOE National Labs to address urgent near-term DUNE computing needs. DOE National Labs in the DUNE collaboration already have very strong teams working on HEP computing. The goal of this effort is to connect these existing teams directly to the DUNE experiment's immediate needs and to build the broader US effort through cooperative work with University groups. We propose a consortium consisting of 5 university groups, partnered with 5 national laboratories. This consortium would be led by PI Heidi Schellman at Oregon State University, with co-PI's at the collaborating institutions.

DUNE Computing Requirements

The overall DUNE computing strategy is firstly to collaborate with the wider HEP community on common solutions and secondly to identify areas where R&D is required to tackle DUNE-specific problems. To that end, the international DUNE computing consortium, which contains a broad spectrum of expertise from several previous and running experiments, has identified three priority areas that require immediate attention to meet DUNE's physics goals:

1) Databases – an integrated database design that will support the full lifecycle of the experiment and the urgent demands during construction.
2) Data handling and computing frameworks - development of a data model, data handling software and software frameworks to deal with the unique DUNE data that spans 6 GB to 460 TB in event size.
3) Collaborative Computing - a standard, portable DUNE Computing environment that can be used by all DUNE institutions, facilitating the work of all collaborators.

Each of these priority areas has been translated into a work package that will combine University groups with National Laboratory expertise. All of these are high-priority, near-term projects necessary to prepare DUNE Computing for running in the next 5 years and the decades beyond. All projects will build on existing expertise and are integrated into immediate DUNE-specific needs. The work packages are:

1. Databases: Construction of the DUNE detector is already underway and analysis of prototype data from CERN is ongoing. This leads to an immediate and pressing need for a coherent design for hardware and calibration databases in the very short term to properly track this construction and testing phase. On a timescale of one year, DUNE will require a coordinated suite of databases with well-designed interfaces to integrate hardware, alignment, calibration, configuration, conditions and data-processing metadata in order to extract optimal performance and efficiently produce physics measurements. Previous experience shows that narrowly-focused, short-term solutions lead to greater long-term maintenance and development costs, poor interoperability and larger operations effort. In cooperation with the existing construction and calibrations efforts of the broader DUNE collaboration, the University of Minnesota, Colorado State University, Argonne, Fermilab and Brookhaven National labs, propose the design and

implementation of the integrated database tools needed for both protoDUNE and the full DUNE experiments.

2. Data handling and frameworks: DUNE's far-detector data consist mainly of Liquid Argon TPC readouts with between 10,000 to 200M samples/channel across up to 1.5 M channels. Data processing includes noise reduction and deconvolution of raw wire signals to produce reduced 2-D hit information. The signal processing that produces 2-D hits is followed by event-level algorithms to reconstruct particle kinematics, differentiate muon from electron neutrino interactions and detect low-energy neutrino signatures from astrophysical sources. Many of the data processing steps appear well suited to new computing architectures and DUNE is actively assessing potential improvements from multi-threading, GPU acceleration and machine learning techniques. Thanks to DUNE's large event size, the layout of data becomes very important as single 6 ms readouts of individual wire planes are close to 15 MB in size. ~~The challenges for machine learning meanwhile lie in the complexity of the models needed to describe the data, which result in large inference times~~. Significant work is needed now to develop the data model and algorithms capable of meeting these challenges in the face of evolving computing systems. They must scale from O(6 ms) single readout planes to O(100 s) supernova readouts and efficiently use a wide range of future computing resources. In addition, novel near detector technologies have been proposed and must be integrated into the overall DUNE software architecture. Fermilab, Brookhaven, Argonne National Lab, Oregon State University and Wichita State participate in this package.

3. DUNE is a large international collaboration with over 180 institutions including 80 in the US. Collectively those institutions have immense personnel and hardware resources that need to work together coherently as a collaboration if the ambitious timeline for DUNE is to be met. Common infrastructure, collaborative tools, code management tools, software tutorials, training, and analysis templates are all crucial to making DUNE data easily accessible to all collaborators. In particular, DUNE requires a build system that creates tested, portable software suites that can be used both by individuals and for large scale data processing. Brookhaven National Laboratory, Fermilab and Oregon State University have expertise in this area and for a small investment could establish a strong backbone for US DUNE collaboration as soon as possible.

The requested funding per institution would be sufficient to support a postdoc or ½ of a senior computing expert working on the project. Costs/per institution range between $120,000 and

$150,000/year. Total cost would be of order ~~1.5 $M per year or 4.5 $M for the full 3 year project.~~ Funded at 1.0 $M/year, 3.0 $M total

<u>Co-PI biographies</u>

Oregon State University – tasks 2,3

Heidi Schellman, Professor - Head DUNE Software and Computing Consortium. Head, Oregon State Dept. of Physics. Formerly computing coordinator for the CCFR, NuTeV, D0 and MINERvA experiments. Former D0 collaboration Luminosity, QCD and EW convener. Design and operation of production systems for Mark II, CCFR, NuTeV, D0 and MINERvA.

Argonne National Laboratory – task 2

Tom LeCompte, Physicist, High Energy Physics Division - Held leadership roles on the CDF, STAR and ATLAS experiments. He has been involved in bringing high performance computing to ATLAS from the earliest days.

Peter van Gemmeren, Principal Computational Scientist, High Energy Physics Division - Manager of Data Management in US-ATLAS Software and Computing operations. Lead Developer of core Input/Output software framework and event data store for the ATLAS experiment.

Brookhaven National Laboratory – task 1,2,3

Paul Laycock, Staff Scientist, Nuclear and Particle Physics Software Group, BNL. US Belle II Computing lead on Data Management and Conditions DB, HSF Data Analysis WG convener. Formerly ATLAS Data Preparation Coordinator, ATLAS Conditions DB coordinator, HSF Conditions DB WG convenor, NA62 offline computing architect, H1 physics board.

Brett Viren, DUNE/MicroBooNE physicist. DUNE FD DAQ lead developer of the Prototype Trigger Message Passing software system, lead developer of Wire-Cell for Liquid Argon TPCs (DUNE and MicroBooNE), Developer and integrator of the Gaudi-based event processing framework for the Daya Bay experiment.

Torre Wenaus, Senior Physicist, Group Leader, Nuclear and Particle Physics Software Group, BNL. US ATLAS HL-LHC Computing manager. Formerly ATLAS Computing Coordinator, ATLAS Distributed Computing Coordinator, HSF Startup Team Co-Lead.

Colorado State University – task 1

Norm Buchanan, Associate Professor, Colorado State University, Lead for DUNE Databases group. Leadership roles on a number of experiments (D0, T2K, NOvA, DUNE), including run coordinator for D0 and T2K, L2 trigger software manager for D0, calorimeter group convener for DZero, PiZero subdetector installation manager for the T2K ND280 near detector, L3 photon detector project manager for LBNE/DUNE, and data-driven trigger group convener for NOvA.

FNAL – tasks 1,2,3

Ken Herner, Applications Physicist II, FNAL Scientific Computing Division. Currently DUNE Production Group coordinator. FabrIc for Frontier Experiments (FIFE) project lead. Production Support Area Coordinator for Open Science Grid. Former co-convener of D0 Higgs group and Trigger group.

Michael Kirby, Staff Scientist, FNAL Scientific Computing Division. Member of MicroBooNE and DUNE. Currently the DUNE Computing Consortium Host Lab Technical Lead. Formerly Data Management and Computing Production Lead for MicroBooNE experiment and FabrIc for Frontier Experiments (FIFE) project lead.

Andrew Norman, Staff Scientist, FNAL Scientific Computing Division, DUNE Computing Architect. HEPCloud project lead. Neutrino science lead for SciDAC-4 HEP Data Analytics on HPCs

Steve Timm, FNAL Scientific Computing Division, Head DUNE Data Management group, HEPCloud Operations lead and Integration/Testing lead.

Tom Junk, Staff Scientist, FNAL Neutrino Division, DUNE software co-manager. Near detector MPD software coordinator. Former DUNE software and computing coordinator.

University of Minnesota – task 1

Marvin Marshak, Professor, Head US DUNE Scientists group, 40 years' experience in the fabrication, installation and operation of large underground detectors. Formerly Director of Soudan Underground and NOvA Ash River Laboratories, Head of University of Minnesota School of Physics and Astronomy and Senior Vice-President of the University. Participated in Sanford Lab Project in various roles since its initial conception in 2000.

Hajime Muramatsu, Senior Research Associate, University of Minnesota. Led development of mobile-device-based APIs for DUNE hardware database at the University of Minnesota. CLEO offline librarian, led absolute energy calibration of crystal calorimeter, co-convener of charmonium working group. Hardware development of BTeV RICH detector. Co-convener of charm physics working group of BESIII, maintained computing farm.

Wichita State University – task 2

Mathew Muether, Assistant Professor – Software Integration Coordinator for the DUNE Near Detector Consortium. Previously, neutrino interaction physics working group convener for the NOvA Neutrino Experiment. Run coordinator, data quality group convener, and production expert for the NOvA Neutrino Experiment.