Summary of Data Management Principles

SuperCDMS SNOLAB Experiment

Experiment description

The SuperCDMS SNOLAB experiment employs cryogenic germanium and silicon crystals to detect and measure the predicted rare scattering of dark matter particles on nuclei. The detectors will measure both ionization and athermal phonon signals from dark matter interactions to achieve a very low recoil energy threshold and exceptional discrimination between a nuclear recoil signal and electron recoil backgrounds.

The SuperCDMS SNOLAB experiment will consist of a mixture of germanium and silicon target detectors, with some of the detectors operating in a high-voltage mode. It will be installed at the SNOLAB underground laboratory near Sudbury, Ontario, Canada. Science data taking is planned to start in 2020.

DOE's roles in the experiments

DOE, together with NSF and the Canadian Foundation for Innovation (CFI) have funded the Project to construct and deliver the SuperCDMS SNOLAB experiment. The SuperCDMS Collaboration will operate the SuperCDMS SNOLAB experiment with support from DOE, NSF and Canadian (CFI and NSERC) funding. The SLAC National Accelerator Laboratory (SLAC) will manage the operation of the SNOLAB experiment, under contract with the DOE. DOE operations funding will support operations staff at FNAL, PNNL and SLAC. DOE operations funding will also cover shift travel for Collaboration members to work onsite at SNOLAB, plus material supply costs for operating the experiment. A small portion of the DOE operations budget will pay for computing and management of the SNOLAB data, which will be hosted at SLAC and FNAL.

DOE operations funding is provided directly to SLAC and FNAL, and all M&S purchases are done through the laboratories' purchasing departments. DOE-supported travel is handled through the SLAC and FNAL travel offices, including for DOE-supported university groups.

Partnerships

The DOE Office of High Energy Physics (OHEP) works in close cooperation with NSF and CFI to support the SuperCDMS SNOLAB experiment. SLAC, managed by Stanford University, is the host laboratory for SuperCDMS SNOLAB, and provides management and safety oversight for the SuperCDMS experiment. SNOLAB will provide the underground site and associated infrastructure to support the experiment.

DOE OHEP funding supports the operation of the experiment, for operations-specific DOE lab staff at SLAC, FNAL, and PNNL and support of shift travel for collaboration members at those labs and members from DOE-funded university groups within the Collaboration. The NSF funding supports several of the US universities within the Collaboration and is primarily used for

travel for shift work at SNOLAB, plus early operations activities at the CUTE, NEXUS and TUNL facilities. Likewise CFI and NSERC in Canada support scientists at Canadian universities within the Collaboration. Scientist support is provided through base funding from either DOE, NSF or NSERC at each institution. There are no overarching agreements between the DOE and NSF regarding data management for SuperCDMS.

Organization – Agency/Lab level

SLAC is the host laboratory for SuperCDMS SNOLAB, and is the home institution of the experiment's operations manager, and provides oversight of the experiment operations. The DOE OHEP program manager for the SuperCDMS SNOLAB experiment operations is Kathleen Turner and the NSF program manager is James Whitmore.

Organization – Experiment level

An operations team within the Collaboration operates the SuperCDMS SNOLAB experiment for the Collaboration (Figure 1). The operations team, led by the operations manager, meets weekly to coordinate operations tasks. The operations team and SNOLAB laboratory staff will monitor the cryogenics system, overseen by both operations management and an FNAL engineer. Operations team members will maintain the electronics, data acquisition and computing hardware at SNOLAB. Data acquisition and data quality groups will include scientists from both lab and university groups, and together they will monitor the data being taken.

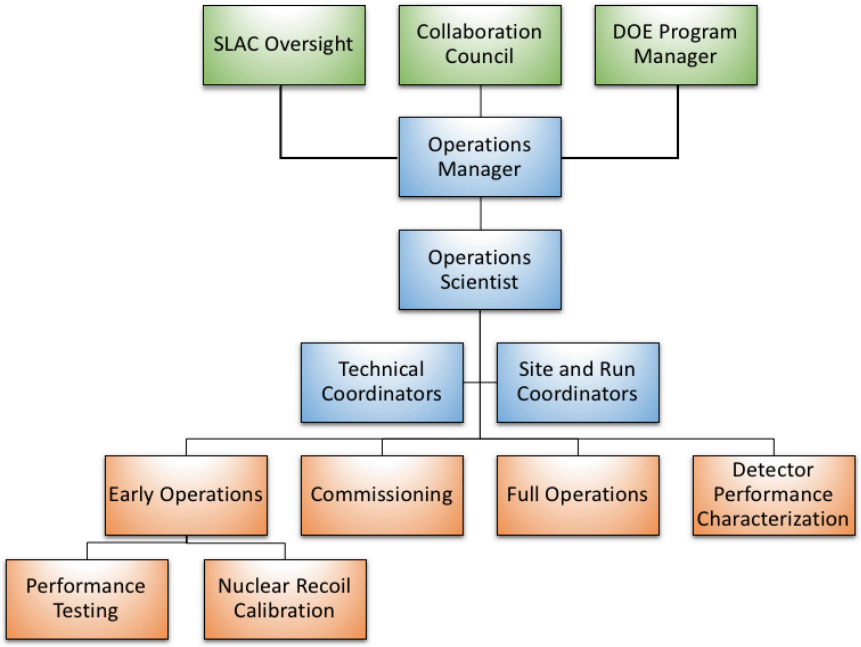

Figure 1: SuperCDMS SNOLAB Experiment Operations organization

Collaboration

The SuperCDMS Collaboration is an international collaboration consisting of approximately 100 scientists from 23 institutions including three national laboratories (FNAL, PNNL, SLAC). Most institutions have base grants funded either by DOE, NSF or CFI and NSERC. The Collaboration Spokesperson is elected from the collaboration and may be from either a university or one of the national laboratories. Day-to-day affairs of the collaboration are managed by the Executive Committee, which reports to the collaboration Council. The Council includes a Council chair elected from the Council, plus the PIs, working group chairs and elected representatives from the student/postdoc committee. Working groups and experiment operations support organization within the Collaboration are shown in Figure 2.

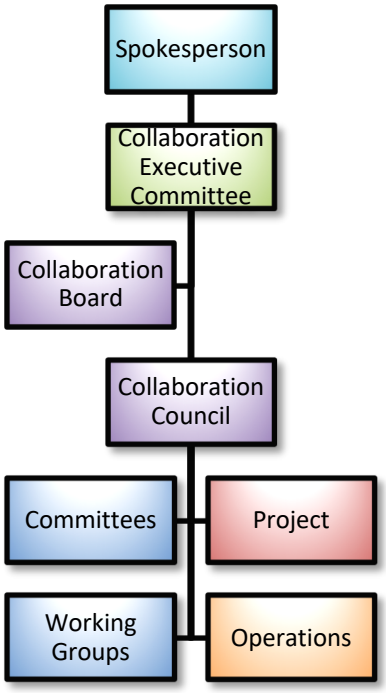

Figure 2: SuperCDMS Collaboration organization

Data policy management

The SuperCDMS Collaboration Council sets the data management policy for the collaboration.

Data Description & Processing

The raw data produced by the SuperCDMS SNOLAB DAQ system will consist of ionization and phonon waveform traces, as well as environmental data such as cryogenic system data and run-time data (e.g. trigger rates and detector state). The raw data are stored in binary files and SQL databases.

There are two main types of data:

- Calibration data using radioactive sources
- Low background data for dark matter search

Most of the data taking will be in the low background mode with an expected trigger rate of 0.03 Hz. A limited amount of calibration data will be taken per week with expected trigger rate of 5 Hz/detector. The data volume will be dominated by the calibration data.

Real-time processing with a limited number of algorithms and preliminary calibrations will be done first at SNOLAB in a Tier 0 computing facility at the surface. These data are used to monitor the stability of the experiment and check the quality of the incoming data in real time. The raw data are then transferred to SLAC and FNAL for processing and long term storage. All the raw data are processed at SLAC on the SLAC computing farm using the full reconstruction package, including blinding of the dark matter (WIMP) search data.

The first level processing produces ROOT files containing variables calculated from pulse reconstruction analyses, such as optimal filter, and environmental data analyses. A second level processing produces the calibrated quantities.

Monte Carlo simulation computing is performed mainly at SLAC and Texas A&M.

After analysis of the WIMP search data and publication of the results, three types of data are produced:

- Candidate data: Information (e.g. charge and phonon energy) about the events passing all the selection criteria.
- Exposures and efficiencies: Final WIMP efficiency for each detector as a function of total phonon energy, after applying all selection criteria, and exposures for each detector.
- Nuclear recoil energy scale: parameters used to calculate the recoil energy of the events.

Data Products and Releases

All data releases will be made publicly available on the collaboration website at SLAC: https://supercdms.slac.stanford.edu/. The data used in a given publication will be made publicly available at the time of publication or shortly thereafter. A document with instructions and detailed descriptions of the data release (including any quality cuts applied, efficiencies, exposures and nuclear energy scale) will be provided for each data release. An email address at which the collaboration can be contacted regarding any questions about the release also will be provided in the documentation.

Plan for Serving Data to the Collaboration and Community

The collaboration is committed to making all experimental data available to collaboration members as quickly as possible. Raw data will be available to view immediately after the data have been collected. Processed data take longer to prepare, but will be made available to the

Collaboration in a timely manner at Tier 1 data centers at SLAC and FNAL. Collaboration members with proper login credentials can obtain/view/analyze data from those locations.

The Collaboration is not planning to release raw or processed data to the community. Nevertheless, the Collaboration will provide data from all finished analyses alongside each specific publication. The final datasets used in publications are typically much smaller in size and will not require special software tools to analyze. The decision to not provide all data to the community was made due to cost benefit considerations. We don't have the resources within our collaboration to provide an easy to use dataset, along with analysis tools. Additional resources would need to be invested to accomplish this task, that we feel are better spent elsewhere.

Plan for Archiving Data

Data collected onsite at SNOLAB are copied to a RAID disk array on the surface and then copied to SLAC and FNAL over a wide-area network. At both SLAC and FNAL, the raw datasets are redundantly archived on a series of RAID disk servers and spooled to tape in a central storage facility. At SLAC, the data are then further processing on SLAC computer farm and then mirrored to FNAL for archiving, and made available to the other institutions for retrieval and analysis. The most recent processed data are stored on the RAID data servers at SLAC and FNAL. Portions of the processed data are also copied to and stored at Tier 2 data centers at other collaborating institutions such as SMU and Texas A&M for science analyses.

Plan for Making Data Used in Publications Available

The collaboration is committed to provide data from all publications to the wider community. We strive to make data relevant to a given publication available at the same time as the publication, but may not achieve that goal in all cases. If we are not able to provide the data from a given publication at the same time as the publication becomes public we will append the publication with the relevant data as soon as possible. Along with the data, we are committed to provide scripts that will show how the data can be used and visualized. Data will be provided in a standard format (e.g. text and/or ROOT files) and any scripts provided will be written in a widely used programming language (e.g. Python). The exact data format and script language will be left to the analysis lead’s discretion.

Responsiveness to SC Statement on Digital Data Management

This data management plan follows SC Statement on Digital Data Management with the exception that the entire data set is not made public. The plan describes our justification for this exception.