Summary of Data Management Principles
SuperCDMS SNOLAB Experiment

Experiment description

The SuperCDMS SNOLAB experiment uses cryogenic germanium and silicon crystals to detect and measure the predicted rare scattering of dark matter particles on atomic 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 with high voltage bias applied across them. It will be installed at the SNOLAB underground laboratory near Sudbury, Ontario, Canada. Science data-taking is planned to start in 2025.

The DOE's roles in the experiment

The DOE, together with the National Science Foundation (NSF) and the Canada 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 funding support from the DOE and the NSF in the USA and the CFI and the Natural Sciences and Engineering Research Council (NSERC) in Canada. The SLAC National Accelerator Laboratory (SLAC) will manage the operation of the SuperCDMS SNOLAB experiment, under contract with the DOE. The DOE operations funding will support operations staff at the DOE's FNAL, PNNL and SLAC laboratories. The DOE operations funding will also cover shift travel for Collaboration members to work onsite at SNOLAB, plus materials and supplies (M&S) costs for operating the experiment. A portion of the DOE operations budget will pay for computing and management of the SuperCDMS SNOLAB data, which will be hosted for the Collaboration at SLAC and FNAL.

DOE operations funding is provided directly to SLAC, and all M&S purchases are done through the DOE laboratories' purchasing departments.

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 operations, 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 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 SuperCDMS Collaboration and is primarily used for travel for work at SNOLAB, plus early operations activities at the Cryogenic Underground TEst Facility (CUTE) at SNOLAB, and at the Northwestern EXperimental Underground Site (NEXUS) test facility at FNAL. Likewise, in Canada, the CFI and NSERC provide funding to 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 and Canada 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 overall management of the experiment operations. The DOE OHEP and NSF Physics each have program managers for the SuperCDMS SNOLAB experiment operations, and together with Canadian participation, and are organized into a Joint Oversight Group (JOG) to oversee the management of experiment operations.

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 run and monitor the experiment, overseen by operations management. Operations team members will maintain the cryogenics, electronics, data acquisition and computing hardware at SNOLAB. Data acquisition and data quality groups will include scientists from both laboratory and university groups, and together they will monitor the data being collected.

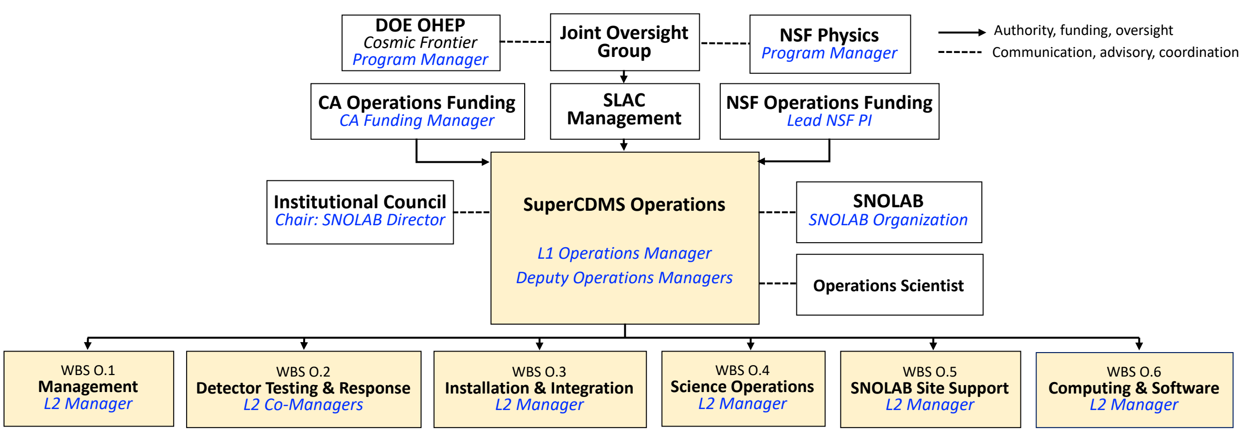

Figure 1: SuperCDMS SNOLAB Experiment Operations organization

Collaboration

The SuperCDMS Collaboration consists of about 130 full members and about 50 associate members from 27 institutions in 6 countries, including three US national laboratories (FNAL, PNNL, SLAC). Most institutions have base grants funded either by DOE, NSF, CFI or NSERC. The Collaboration Spokesperson is elected from the collaboration and may be from either a university or a national laboratory. Day-to-day affairs of the collaboration are overseen by the Executive Committee, which is led by an elected Spokesperson from within the collaboration. The Executive Committee reports to the collaboration Council, which is the decision-making body within the Collaboration, composed of the Collaboration members. The Council includes a Council chair elected from the Council, plus the PI scientists, and working group chairs and elected representatives from the student/postdoc committee. Committees and working groups within the Collaboration and experiment support organizations are shown in Figure 2.

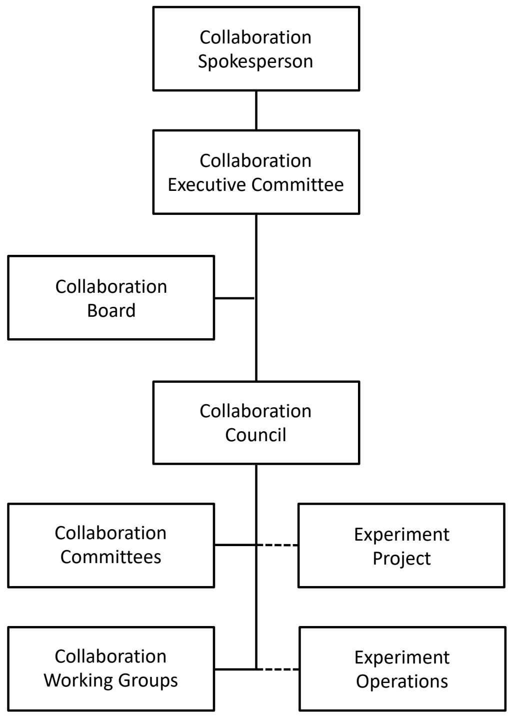

Figure 2: SuperCDMS Collaboration organization

Data policy management

The SuperCDMS Collaboration Council sets the data management policy for the collaboration. Collaboration policy documents are available through https://supercdms.slac.stanford.edu/, the home website of the Collaboration.

Data Description & Processing

Generated research data generally fall into categories of experimental data, simulation data, or analysis results. Experimental data include (but are not limited to) instrument-specific output that will be stored in raw form or in compressed formats using appropriate data compression algorithms, and instrument metadata that define experimental conditions, as well as a description of algorithms and processing used to transform the raw data. Simulation and analysis products include computation output, including (but not limited to) Monte Carlo simulations as well as tools that process and analyze data and simulation results.

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 collected data: calibration data using radioactive sources, and low background data for dark matter searches. Most data will be acquired in the low-background mode, with a low trigger rate. However, less frequent acquisition of calibration data – taken with a higher trigger rate – is expected to dominate the total data volume.

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 software package, including blinding of the dark matter search data. The first level processing produces ROOT files containing variables calculated from pulse reconstruction analyses, such as an 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 University.

Analysis for dark matter generally results in three types of data:

- Candidate data: information (e.g., charge and phonon energy) about the events passing all the selection criteria.
- Exposures and efficiencies: dark matter detection efficiencies (after applying all selection criteria) for each detector as a function of total phonon energy, and exposures for each detector.
- 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/public-data. The data used in a published paper 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 energy scale parameters) 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 collection. 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, and 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. The Collaboration does not have the resources to provide a general easy-to-use dataset, along with analysis tools. Additional resources would need to be invested to accomplish this task, which the Collaboration has decided would be better spent elsewhere.

Plan for Archiving Data

Data collected onsite at SNOLAB or FNAL are first copied to redundant disk arrays and then copied to SLAC and FNAL servers over a wide-area network. At both SLAC and FNAL, the raw datasets are redundantly archived on RAID disk servers and copied to tape in a central storage facility. At SLAC, the data are then further processed on the SLAC computer farm and made available to the other Collaboration institutions for retrieval and analysis. The raw and processed data are mirrored to FNAL for archiving, and available for data recovery in case of data loss. The most recent processed data are stored on the data servers at SLAC. Portions of the processed data are also copied to and stored at Tier 2 data centers at other collaborating institutions, such as the Digital Research Alliance of Canada (DRAC) 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, then 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 leader's discretion.

Responsiveness to SC Statement on Digital Data Management

This data management plan follows the 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.