# Elements: Improving tools based on data-description standards for gigabyte-scale data sets

March 31, 2019

Amy Roberts

*University of Colorado Denver*

# Project Summary


## Elements: Improving tools based on data-description standards for gigabyte-scale data sets

**Overview:** The PI proposes to improve a set of data-analysis tools that require a description of the data format - rather than requiring a particular data format - to handle the gigabyte-scale datasets common in nuclear physics. The long-term goal of this work is to increase the accessibility of science analysis for both active researchers and science students.

In pursuit of this goal, this proposal has three objectives: (1) Improve existing software that provides access to data in any user-described format so that it works well with gigabyte-scale datasets. (2) Build community adoption and support of these tools. (3) Increase access to analysis of scientific data sets for undergraduate students, early-career researchers, and scientists who do not have access to dedicated software support. These objectives will be met as follows:

*(1) Improving standards-based analysis software:* Scientists who wish to analyze custom-format data can already do so with existing open-source software, e.g. kaitai-struct. However, this software has poor performance for files larger than a gigabyte, limiting its usefulness to the nuclear physics community. The PI proposes to improve this software to allow responsive analysis of gigabyte-scale datasets in python. This work is feasible thanks to libraries built and maintained by the IRIS-HEP collaboration that provide an intuitive interface to data structures optimized for rapid access to large, event-based data sets.

*(2) Building community:* The PI proposes yearly workshops to bring together developers and scientists for training, user feedback, and development of this software.

*(3) Increasing Access:* To fully participate in nuclear physics research, students must have extensive scientific computing skills. Mentor networks help some students through this maze but are not equally available. The PI requests funding for undergraduate students to develop and test documentation and training materials for the standards-based analysis tools and for fundamental scientific computing concepts.

**Intellectual merit**: This work supports the science effort of the Super Cryogenic Dark Matter Search, an experiment that seeks to better understand the nature of dark matter. This work supports additional high-priority science in the nuclear physics community such as the origin of the elements in the universe, fundamental symmetry testing, and radiation monitoring for national security.

**Broader impacts**: The PI proposes this work because it would be immediately useful to her own work on dark matter and because software that can provide easy access to any user-described data format would meet an urgent need of communities dealing with gigabyte-scale data. The PI requests funds to develop and field-test documentation for this software to maximize its usefulness to the community.

Right now, the necessity for local development of analysis code restricts participation in science analysis. With a small set of well-documented community tools based on data-description standards, along with open-access material introducing the computing concepts required to use those tools, the PI hopes to increase research access to a much wider group of individuals and increase the time expert scientists can spend doing science.

## Contents

# 1 Introduction

## 1.1 The Problem

The PI proposes improvement and community building for a common set of tools to make it easier for scientists to access their data.

In nuclear physics, in neuroscience, in remote sensing applications, custom data formats are commonplace both for existing and newly-recording datasets. This fractures the software ecosystem. Even though the nuclear physics community performs largely similar analyses on largely similar datasets, groups typically develop their software locally. Poor documentation of these isolated code bases can have real, negative impact on science progress.

**S has a data set that would be valuable to include in an analysis** that combines different types of carbon scattering data. Because the data was taken with a polarized beam, it would provide a valuable constraint to her global fit. But she can't include it because the only code that can read the data won't compile on her machine. She's not familiar enough with Fortran to fix it, the original author died a few years ago, and the beamline has been decommissioned. She could still write a paper, but the results aren't all that interesting without this data.

**P would like to add an additional detector to his setup** – this would allow him to set a stronger limit on a reaction rate that effects how stars create heavy elements. A colleague lent him a digitizer to instrument his extra detector, but the analysis code he has is for CAEN instruments and the loaner digitizer is from XIA. After burning a weekend trying and failing to adapt his code to handle the XIA data, he decides to try the experiment without the extra detector. He might get a useful constraint on the integrated cross section, and maybe if the funding for his postdoc comes through they'll have time to sort out the code.

**All Q wants to do is look at the data from a detector used to search for dark matter.** He got one of the files onto his computer, but when he tried to open it in Word it looked like a bunch of garbage symbols. Three frustrating days later, he now knows the data is "binary" and he's downloaded "source code" that should allow him to look at the data. When none of the commands in the readme file make sense, he asks his professor for help, who tells him to talk to the postdoc, who apparently doesn't respond to emails. It takes almost an entire summer, but Q does eventually figure out how to compile the code, run the code, and look at a single event.

## 1.2 Building a Solution

One way to solve this problem is to define a standard way to describe a data format. This allows development of tools that provide access to that data based on its description. Scientists can then keep their custom-format data and âĂŞ by providing a description of that data âĂŞ gain access to analysis tools that are supported by a broader community. The work I propose here works towards this solution in two ways:

The long-term goal of this work is to build and create community buy-in for a set of tools for the nuclear physics community to access binary data. The intent is to address one piece of a common software need in the community that is repeatedly solved in isolation and that costs the entire community unnecessary time and limits access to data and science. In pursuit of this goal, this proposal has four objectives: (1) Define a data-description language that allows scientists to describe any binary data format. (2) Improve existing software that automatically builds a convenient data-access library for any user-described data so that it works well with gigabyte-scale datasets. (3) Build community and support adoption of these community tools. (4) Increase access to analysis of original scientific data sets for undergraduate students, graduate students and postdocs, early career researchers, and scientists whose labs do not have access to dedicated software support. These objectives will be met as follows:

*(1) Defining a data-description standard:* The PI will use two already-existing standards that both have broad user support, DFDL and Kaitai-Struct. Kaitai-Struct will be used and where necessary further defined as the software created around this standard is better-suited to the needs of data analysis.

*(2) Improving standards-based analysis software:* Scientists who wish to generate a convenient library that allows them to read custom-format data can already do so using the existing Kaitai software. However, this software has poor performance in python for files larger than a gigabyte and this significantly limits the usefulness of this software to the nuclear physics community, particularly as data acquisition systems move to storing digitized pulses rather than pre-processed data. The PI proposes to create another Kaitai target that allows scientists to auto-generate a library that targets the awkward-array interface built and maintained by the IRIS-HEP collaboration. This data structure is designed to give rapid data-analysis access to the high energy physics community and is routinely tested on data sets that exceed several hundred gigabytes.

*(3) Building community:* The PI proposes to hold yearly workshops to bring together developers and scientists for training, user feedback, and software development. In addition, the PI has requested funds for graduate and undergraduate stipends and travel awards to foster community participation.

*(4) Increasing Access:* Many students who have the interest and ability to participate in science analysis are blocked by a lack of training in scientific computing. The often-complex scientific computing ecosystem is poorly represented in most physics curriculum and even computer science courses do not usually prepare students for the variety of systems they need to navigate to do typical data analysis. The PI requests funding for undergraduate students to develop and test documentation and training materials for fundamental scientific computing concepts and skills in addition to testing the documentation for the standards-based analysis tools.

### 1.3 Long-term research goals

### 1.4 The Team

**This paragraph needs to be added after we do decide on the sections** This project description is organized as follows. Sections **??**–**??** summarize the science and hardware of the SuperCDMS SNOLAB experiment. Section **??** describes the underground facilities to be used in the pre-operations portion of this proposal. Section **??** describes the plan of work of this effort, and Section 6 describes the schedule. Section **??** gives an overview of the management for this proposed grant, and Section 7 describes the broader impacts of the work. Finally, Section 8 summarizes the results from previous support.

## 2 Plan of Work

### 2.1 Introduction

### 2.2 Science-driven

This project serves the immediate needs of researchers in the dark matter community and the experimental nuclear physics community by providing a common toolset for analyzing data in any format.

The PI expects that

- Multiple research projects across the NSF directorate will be more productive because they can use existing, documented tools rather than building their own. The PI intends to estimate this impact with citations from scientific papers.
- Increased involvement of undergraduate researchers in science analysis due to improved documentation and an extended support network. The PI intends to measure this through undergraduate involvement in her own lab, community surveys, and tracking community forum data.
- Several example analyses will be publicly released, with accompanying documentation and support information for pre-requisite computing skills. The intent is for these educational materials to be accessible to someone with no domain knowledge. The PI believes these training materials will be an equitable training resource.

### 2.3 Innovation

A common limitation of data-analysis software that is entirely home-grown is that it does not scale as data grows and changes - a human has to update or rewrite the code if the requirements change substantially.

This library makes heavy use of existing libraries. The benefit is that as those libraries improve and scale to larger data sets, this software inherits that improvement.

In both cases, significant human effort is needed to adapt the software to changing data and needs. But by leveraging well-supported, open-source libraries, the burden is shifted away from an individual scientist and towards an active community that is highly motivated to solve similar problems. This library serves as sugar to allow scientists with many different data formats to take advantage of these popular libraries.

The danger to this approach is the same - if these libraries lose community support or focus on very different problems then this library will lose relevance over time. To mitigate this risk, the PI is focusing on integrating with a library supported by the IRIS-HEP collaboration and with pandas, which enjoys extreme popularity in the data science community.

### 2.4 Close collaboration among stakeholders

The PI proposes to engage both cyberinfrastructure experts and the experimental nuclear physics community by (1) working closely with pilot experiments to build software that works effectively for scientists analyzing event-based data, (2) holding yearly workshops intended to foster interaction between the scientists using the software and cyberinfrastructure developers, and (3) attending conferences that will allow outreach to the scientific community (for example, the Low Energy Community Meeting) and the cyberinfrastructure community (for example, CHEP).

The PI has working relationships with scientists in the SuperCDMS collaboration and the XIA corporation, both of which are interested in exploring the proposed software as solutions to analysis needs.

In addition, the IRIS-HEP collaboration is interested in this work as it would extend their awkward-array library to a broader audience. Awkward array was developed by Jim as part of DIANA/HEP (http://diana-hep.org, OAC-1450377) and that effort will continue with IRIS-HEP (http://iris-hep.org, OAC-1836650). Collaborating with IRIS-HEP gives us access to experienced cyberinfrastructure developers who have focused on developing software suitable for terabyte-scale data.

### 2.5 Building on existing, recognized capabilities

The proposed work builds on existing capabilities and communities in several ways:

- **The PI proposes to use already-existing data description languages.** The languages Katai Struct and the Data Format Description Langauage both have active communities and tools that work with data when provided a description. Kaitai Struct is the target for the proposed work because (1) it is more human-readable than than the XML-based DFDL, and (2) Katai Struct generates code libraries that allow users to load their data into the programming environment of their choice; DFDL currently works by providing an XML or JSON equivalent of the binary data. While this is a powerful approach because any language with an XML or JSON parser can now read the data, it also produces a secondary data file that is an order of magnitude larger than most binary files. This makes DFDL, in its current state, unusable for scientists with gigabyte-scale data sets as it would make the required storage space for analysis prohibitively expensive.

- **The PI proposes to use already-existing infrastructure for the data-analysis library.** Scientists who would like to avoid writing custom software to read their binary data can already use the Kaitai Struct compiler to generate libraries to read their data in python, C++, and a multitude of other languages. The advantage is that there is substantial support documentation and an active community available for troubleshooting. The disadvantage is that the current Kaitai Struct python compiler stores the data in a structure that does not provide adequate speed performance for gigabyte-scale data sets. By improving the existing Katai Struct compiler software, we can build a science-ready analysis library and scientists can benefit from the existing community support and documentation.

- **Use a supported and optimized data structure** for the improvements to the Kaitai Struct compiler. The "awkward-array" library was developed by DIANA/HEP and is now supported by IRIS-HEP and is part of a set of libraries designed to provide flexible data-analysis tools for the high-energy physics community. The awkward-array data structure is optimized for fast queries on an event-based data set and as such is ideal for the majority of nuclear physics data. By choosing this data structure as the target, we bring the optimized and convenient analysis environment of awkward-array to any scientist who describes their data with the Katai Struct language.

- **Provide analysis tools for the python environment and training materials that take advantage of the python ecosystem.** Python is a popular analysis environment in the field of big-data and has enjoyed significant adoption in the scientific community; enough so that python support is compiled in the dominant high-energy physics software, ROOT, by default. By providing a python library for data analysis, scientists can make use of a full ecosystem that supports data analysis: numpy for convenient array manipulation; scipy for fitting; matplotlib for producing publication-quality figures; and even numba for easy compilation of code that needs to run fast. This entire environment is easily installed - even for users without administration privileges - through the Anaconda Python distribution. There are many free and paid programming environments that are availble, notably the Jupyter environment. Code written in this environment is particularly nice as a tutorial because it is rendered nicely on github, gitlab, and interactive notebooks can be opened in one click through binder. By providing a small set of introductory documentation, scientists can benefit from the effort the python community has put in to lower the barrier for use.

### 2.6 Project plans, and system and process architecture

The timing of the proposed work is driven by the proposed, yearly workshops that focus on (1) teaching scientists how to use the tools to access their data, (2) working with scientists to perform their analyses in the python environment, (3) identifying improvements needed for the software to be easy to learn and useful in analysis, and (4) bringing developers into close contact with the science community using their tools. Each workshop will result in an updated roadmap for the software.

Thus, the workshops - and software releases that include testing, documentation, and example analyses - are the primary milestones of the proposed work.

The work for each yearly cycle can be broken down into the following categories: development of basic computing skills learning material; development of the data-access library; planning and execution of the workshop; and a community-driven update of the roadmap. See Table 2.6 for details on who will perform this work.

The minimum requirements of the work determine the work plan and are the following:

1. If students or staff move on to other positions, their replacements should be able to get up to speed in a month or less.
2. Someone with no domain knowledge but reasonable persistence should be able to run the example analysis within a week.
3. Someone with no domain knowledge but reasonable persistence should be able to analyze their own data within a month.
4. A scientist who uses the access-data library to obtain a science result should know how to cite the software.
5. A scientist experiencing trouble using the software should be able to determine how to get help quickly (within five minutes of searching).
6. A scientist who wishes to improve the code should be able to quickly determine how to contact the developers and how to change, test, and push the code.

The scope of the proposed software is relatively modest: copy an existing framework and adapt it so that it stores data in awkward-array structures rather than slower, dictionary structures. The development and testing of this code will take time - but reference code for similar work exists, there is robust community support, and there is a developer guide that gives specific instructions for developers who wish to extend the existing Kaitai Struct code in this way. The proposed work is feasible because it connects two libraries that are both designed for this purpose.

The majority of the proposed work is in making this software easy to use for scientists, and making it easy for the community to participate in the direction and development of the software. This requires robust documentation for both users and developers. Users will require installation instructions, instructions for using the library, and guidance on how to adapt the examples for their own analysis needs. Developers will need additional documentation: instructions for changing the code and testing the code, and instructions and guidance on contributing their changes to the project.

To meet the minimum requirements, the workplan involves creating of initial documentation and an automated testing suite by the Professional Research Associate and example analysis created by the Master's student.

Undergraduates will begin either by working on new scientific computation skills or improving or adding to material of already-developed scientific computation skills. Students who join the lab currently have two first projects to chose from: working through an introductory lab on water simulation, or working through an example analysis of gamma-spectroscopy data. The students try to perform their work using existing documentation; the PI provides guidance when this is inadequate. This provides an opportunity for students to design improvements and learn the basics of contributing to a code repository. In addition to providing valuable training for the students, this process identifies gaps in the documentation that are often invisible to experts.

This initial work is expected to result in improvements and additions to web-accessible tutorial materials. In addition, this training will provide a foundation that will allow the students to attempt the following actions:

- Successfully follow a simple example analysis using the data-access library.
- Successfully follow a tutorial to make and share a change to the library documentation.
- Successfully follow a tutorial to make, test, and share a change to the library source code.

| Work | Who | Notes |
|---|---|---|
| Roadmap & workplan development | PI, PRA, Master's student | |
| Basic skills materials | Undergrads | |
| Data-access library development | | |
| code dev | Professional Research Assistant | this includes testing, user documentation, and dev documentation |
| contributing | Professional Research Assistant | this includes automated testing and contribution guidelines and instructions |
| example analyses | Master's student | |
| documentation testing | Undergraduates, Master's student | documentation testing may feed back into additional skill documetation |
| Workshop | | |
| Recruiting | PI | |
| Organization | Professional Research Assistant, Undergraduates | |
| Pre-workshop analysis coordination | PI, Master's student, PRA | |
| Community-driven update of roadmap | PI, PRA, Community | |

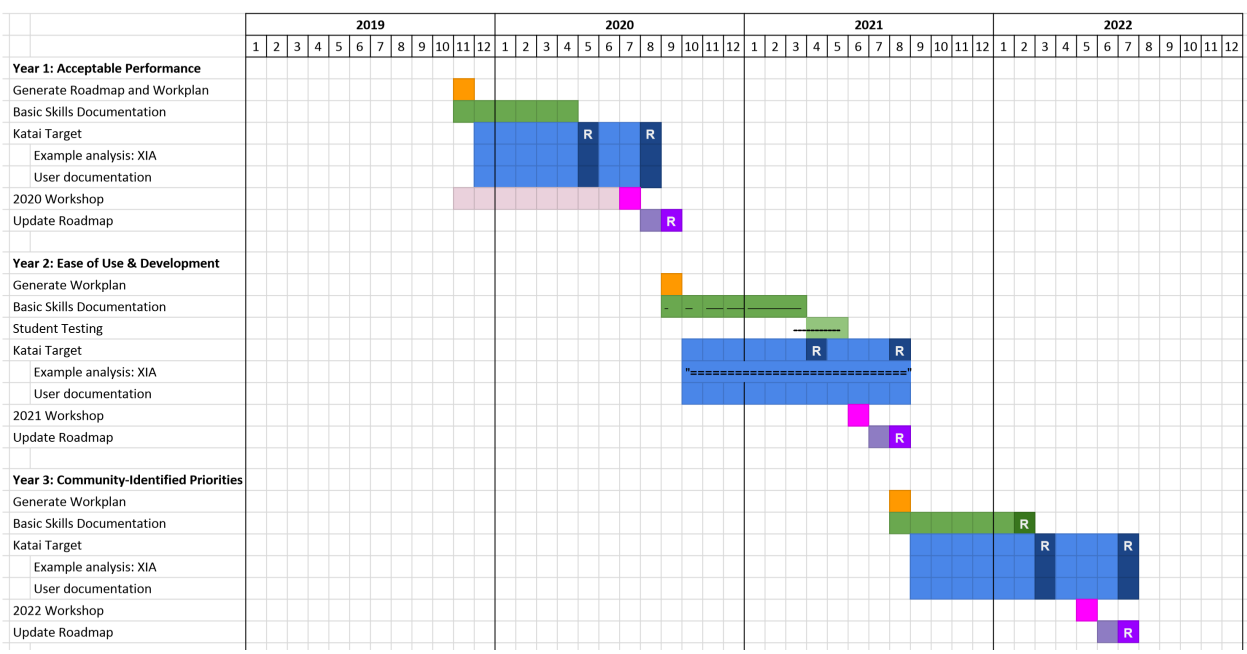


**Figure 1:** Schedule for the proposed work.

**Architecture of the software:** The architecture of the Kaitai Struct compiler that targets an awkward-array data structure will follow that of the existing Katai Struct software. Implementing a Kaitai Struct compiler for a new language requires

1. Writing a "runtime library" that provides a standard stream interface in the target language. For example, one of the functions every language needs to have defined is a method that returns the size of the file or string stream. By writing a "size" function for the language of interest that follows the Katai Struct API, the code generation becomes simpler.
2. Writing a "compiler" that translates Kaitai Struct concepts, implemented in Scala, into the target language.
3. Writing a test runner for the new language.

The proposed work targets the python environment, and there is already a Kaitai Struct python compiler. The "compiler" for the python implementation, however, stores data in native-python data structures that provide inconveniently slow access to standard queries on large, gigabyte-scale data sets.

However, the changes that need to be implemented to instead store the data in the faster awkward-array data structure are restricted to the compiler code. The runtime library provides a convenient interface for reading data from a file or stream - this code only cares about the file system interface and does not need to change. The python test interface will need to be updated as the access syntax for the data will change slightly.

Although there is opportunity for improving the speed of the data load, this development will instead focus on adhering to the existing format and style of Kaitai Struct. The goal is to make the existing Kaitai Struct community useful to scientists who work with gigabyte-scale data sets; waiting for a few minutes for the data to load is not ideal but is typical of many locally-built solutions. We can address the more-critical issue of rapid data queries while staying well within the existing framework of Kaitai Struct and intend to do so for the initial implementation of the software.

If we find that data-load times are a significant issue for the nuclear physics community then we will consider more substantial changes to the Kaitai Struct compiler and runtime libary.

```
def size(self):
    # Python has no internal File object API function to get
    # current file / StringIO size, thus we use the following
    # trick.
    io = self._io
    # Remember our current position
    cur_pos = io.tell()
    # Seek to the end of the File object
    io.seek(0, SEEK_END)
    # Remember position, which is equal to the full length
    full_size = io.tell()
    # Seek back to the current position
    io.seek(cur_pos)
return full_size
```

```
uint64_t kaitai::kstream::size() {
    std::iostream::pos_type cur_pos = m_io->tellg();
    m_io->seekg(0, std::ios::end);
    std::iostream::pos_type len = m_io->tellg();
    m_io->seekg(cur_pos);
    return len;
}
\caption{The details of this code are not important. What is significant is the name of this fu
The Katai Struct compiler for both C++ and python can use the ‘‘size()’’ function rather than
```

**Architecture of the user documentation:** User documentation should make it possible for users with little to no domain knowledge to use the data-access library for science. Documentation for the use of the library will be stored as text files in the repository with the code. The files will be written in markdown syntax to improve their readability; this will also render them nicely on cloud-based repository hosts such as github and gitlab. The following documentation will be provided:

1. How to get help with questions or issues about the library.
2. How to install the library and its dependencies.
3. An overview explaining what the user will need to provide (data and a description of the data) and what the library will provide (software to read that data).
4. A tutorial walking through the use-case of a scientist looking at simple data with a custom format.
5. Links to additional resources detailing more complex data formats and more complex analyses.
6. Citation guidelines.

**Architecture of the developer documentation.** Documentation intended to facilitate development of the code will be stored in the repository alongside the code. Text files referenced in the top-level README file will detail, for every repository,

1. How to install, develop, and test the code for individuals who wish to make changes.
2. How to contribute changes back to the project. This will provide instructions on the version control practices used by the repository maintainers and instructions for implementing the tests required for changes to be considered for merging with the main code base.

**Architecture of the basic scientific computing skills documentation:** Documentation of basic computational skills and concepts will have several possible forms: (1) Text and images, (2) tutorial videos, (3) jupyter notebooks, (4) printable images that illustrate a focused concept, and (5) links to recommended resources such as Software Carpentry tutorials.

All materials will be licensed with a permissive, open-source license such as CC-BY or MIT. The source for all the materials will be publicly available through a public host such as github or gitlab and will be archived on a content-tracker such as the Open Science Framework or Figshare. Videos will be released on YouTube and licensed CC-BY.

All materials will be disseminated using a static site generated by Antora. Antora is specifically designed for documentation and allows a user to specify a set of repositories containing text files formatted in the Asciidoc markdown language to build a single, searchable documentation site. Because Antora generates a static site, free hosting services are readily available. This solution allows my students to focus on creating material to explain core concepts and practice interacting with version control rather than spending time wrestling with web development.

The topics students choose to document are largely student-led, with some guidance from the PI. Spring 2019 marks the inception of this project, and the concepts chosen by students for illustration have focused on (1) tutorial-format guide for installing python and running a basic python-based analysis of gamma spectroscopy data, (2) instructions for using a docker container to simplify installation of a complex software environment, and (3) a poster explaining what an executable file is.

Documentation that will be provided in this format alongside the scientific computing resources will include

1. instructions on where to get help with the material and how to provide feedback and and file bug reports
2. instructions for those who wish to contribute to the documentation
3. instructions for the deployment of the documentation

**Engineering processes:** A primary goal of the proposed work is to build software that can be supported and maintained by the community. A primary risk of the proposed work is staff turnover and associated loss of knowledge and onboarding time.

The software design, development, documentation, and testing work together to make it easy for the community to contribute to the software development - a goal that mitigates turnover risk as well.

The design process of the software will start with project documentation that describes (1) use cases, (2) requirements, (3) assumptions, (4) key decisions, and (5) definitions. Such documentation is particularly useful for programmers who lack experience in experimental nuclear physics data analysis and makes it easier for skilled experts to contribute to the project. This documentation also serves as a way to focus community discussions into defining a minimum useful scope for the software.

The development of all software in the PI's lab is done using version control software (git) and a central "repository server" that everyone interacts with. Cloud-based servers such as github and gitlab are used because they are easy to use, provide robust backups, and also serve as a platform for dissemination and collaboration.

The PI's approach to version control and software releases prioritizes (1) easy-to-get, working code and (2) rapid updates. This is implemented by building end-to-end tests and configuring automatic test running triggered by any changes to the repository. Rather than insisting on a specific release cycle, developers are encouraged to put their changes on the public, master branch if their code is non-breaking. Code on the public, master branch MUST pass all tests. Semantic versioning will be used to alert users to breaking changes.

Work that breaks tests MUST be maintained on a separate “branch” that is publicly available but that will only be available to users by explicit action. Instructions for developing code on such a branch will be included in the contributions documentation.

The key to this type of development is building simple end-to-end tests, implementing an automated testing framework, and investing heavily in documentation and testing of that documentation up front. This development strategy works well for the proposed project because the software goal is already well-defined and the initial plan for implementation - leverage the existing Kaitai Struct framework as much as possible - is clear. In addition, this development strategy is well-supported by community solutions and the organization of the PI’s lab:

- end-to-end testing of python code - and even tutorial notebooks in jupyterlab format - enjoy a thriving ecosystem in Python.
- many automated testing frameworks are designed specifically to support developers using cloud-based repository hosts such as github and gitlab; many are freely available to open-source projects.
- novices are always on hand to test documentation. The PI maintains a group of approximately six students, many of whom have minimal experience with scientific computing. Giving them goals such as: work through this example analysis and obtain a similar plot exposes conceptual gaps and problems with the documentation while providing excellent training for the student. In addition, students have responded well to the opportunity to make a substantive contribution to the lab.

In summary, the proposed code development will be strongly tied to documentation, automated testing, and documentation testing.

### 2.6.1 Trustworthiness

End-to-end tests of the software will include tests where the outcome is known.

### 2.6.2 Provenance

All releases will be archived with their own DOIs on Zenodo. Guidance to cite the version of the software used will be included in the top-level README of all releases and posted on all related websites.

### 2.6.3 Reproducibility

End-to-end tests of the software will completely define all inputs, including a reference data set, and compare the output to reference products such as histograms. The PI acknowledges that this is in no way a complete test of reproducibility but feels that it will be a useful starting point.

Instructions for system setup will be included in the documentation. Full or partial system specifications are required for automatic test running and will be versioned together with the rest of the code. The most popular of these, Docker, will be used. More complete system specifications as provided by Nix and Guix will be considered if need arises for more complete reproducibility.

#### 2.6.4 Usability

For the proposed software, usability consists of several scenarios:

| Can a scientist easily use this code to do data analysis on a custom-format data set? | |
|---|---|
| Is the scientist aware that this software exists? Can the scientist find the software easily even if all they recall is a vague description? | Publications citing DOIs hosted on Zenodo with published preprints; well-indexed project website; open development on github, gitlab; indexed on python library repositories where appropriate |
| Can the scientist install the software on their analysis computer easily, with or without root access? | Installation documentation; testing of installation documentation by novices; review of systems that need installation support at workshops |
| Does the software apply itself well to this particular analysis need? Can the scientist see how to use the software in their analysis? | Example analyses; testing of example analyses by novices; creation of a "now you try" document to test effectiveness. |
| Does the scientist know where to get help or discuss issues with the software? | All documentation and code will contain a header directing users to the project forum and repository issue tracking. |

| Can an interested individual contribute to the development of the software easily? | |
|---|---|
| Is there a clear description of the requirements and purpose of the software so that developers can decide if they'd like to participate? | Requirements documentation and a description of use cases will precede all programming work and will be versioned with the rest of the code. |
| Is it clear where to get help or discuss the code? | All documentation will link to the forum and the repository issue tracker. The forum will be configured for web crawlers to maximize its discoverability. |
| Are there clear and complete instructions on testing the software locally? | Local and remote testing infrastructure is as high in priority as the initial development of the code; documentation will be written as part of the first efforts. Students will test these local development instructions. Instructions may reference the scientific computing documentation. |
| Are there clear and complete contribution instructions? | Contribution instructions will be developed with high priority once there is an initial passing test, even before there is functioning code. Students will test these contribution instructions. Instructions may reference the scientific computing documentation. |

### 2.6.5 Adaptability

The proposed software intends to support scientific analysis of gigabyte-scale data for the coming decade.

The benefits of the proposed software, compared to the current, group-driven methods, are increased access to scientific analysis through ease of use, quality documentation, and community support. And improved return on invested maintenance time, since effort on a common set of tools can benefit many scientists.

Another advantage of the proposed software is that it leverages existing, well-supported projects to deliver science-ready software. Adaptations to changing data needs and opportunities can potentially come from efforts outside nuclear physics.

There is always the possibility that these dependencies could be abandoned. Archives of all dependent software will be made as a safeguard against this. Other risk mitigation the PI will pursue is significant investment in automated testing and documentation, particularly interface documentation.

## 2.7 Deployment and user outreach:

A key component of the user outreach will consist of annual workshops designed to promote hands-on use of the software and close collaboration with the software developers. To maximize the effectiveness of these workshops,

- Participants will be contacted in advance to begin early coordination of the analysis they're interested in doing with the data-access library
- Communication before and after the workshop will be encouraged through the maintenance of an open forum
- Prototype software will be released along with installation documentation, an example analysis, and contribution documentation prior to the workshop and tested by novices
- Discussion of the community roadmap will be integrated into the workshop and a new release of the roadmap will follow each workshop
- The conference will be registered on the Open Science Framework, providing an archived record of material prepared by participants.

Deployment of this software will use common open-source channels such as github, gitlab, and python package indexes such as PyPI and Conda. Github and Gitlab both provide static page serving for projects.

Project communication will use the built-in issue tracking provided by repository hosts and open forum software such as Discourse. Live-chat is not an anticipated need, but if it becomes clear this would be useful then the PI will consider using freely-available chat services such as Slack, Zulip, or Gitter.

Deployment of the software will also happen through academic channels such as preprint servers, project releases on the Open Science Framework, and/or Figshare.

### 2.8 Acceptance and evaluation:

Community adoption of this software is central to the success and broader impact of this project.

The simplest evaluation metrics of community adoption and use will be (1) citations of software in peer-reviewed scientific papers and (2) the number of contributions to the code from developers outside the PI's group.

Additional metrics that may provide useful information could include: (a) number of downloads from the Zenodo or gitlab site and (b) quantity of interactions on the issue tracker and forum.

The PI intends to use interviews to understand how the software meets (or not) the needs of the community. The yearly workshops will provide an ideal setting for such discussions. In addition, a standing feedback survey will be linked from the project page to capture responses from the people who have the time and willingness to share.

This feedback will directly inform decisions about priorities and have an official outlet in the release of an annual roadmap. Discussion and contribution to the roadmap will be open to the community and will begin at the yearly workshop.

## 3 Deliverables

## 4 Metrics

## 5 Sustained and sustainable impacts

## 6 Schedule

Figure 2 gives a broad view of the activities of the SuperCDMS collaboration centered on the SuperCDMS SNOLAB experiment, including the DOE/NSF G2 Project, the pre-operations and commissioning period covered by this proposal, the actual operation of SuperCDMS SNOLAB (also detailed in the Experimental Operations Plan), and the R&D program that will naturally lead to upgrades for SuperCDMS SNOLAB that allow it to probe the low-mass dark matter region down to the neutrino floor.

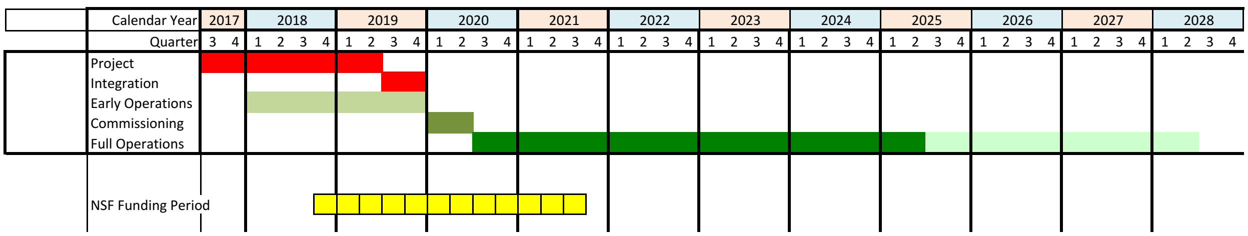


**Figure 2:** Schedule for the SuperCDMS SNOLAB Experiment.

**Year 1: 7/2018–6/2019**
During the first year the focus will be on calibration of the small HV test detectors from Stanford University with the Northwestern ADR at TUNL. These detectors will then be taken to NEXUS, to perform cross-calibration measurements in preparation for the large detector calibrations. The first pre-production detectors arrive at CUTE in January 2018, and the performance and background measurement campaign begins soon after.

**Year 2: 7/2019–6/2020**
The second year is dominated by runs of pre-production detectors at NEXUS for calibration and at CUTE for performance testing and background measurements. The first production SuperCDMS SNOLAB HV tower arrives at CUTE in 12/2018, with testing commencing in 1/2019. Performance and run optimization of all six detectors of the HV tower will continue until 6/2020.

**Year 3: 7/2020–6/2021**
There is no activity planned for NEXUS on this grant in the third year. At CUTE, the SuperCDMS SNOLAB HV tower science run will be taking data until 11/2019, when the tower will be taken out of CUTE to prepare it for installation in the SuperCDMS SNOLAB experiment. Commissioning activities for SuperCDMS SNOLAB will be the main focus of this grant in the last six months of the award, along with the collaboration-wide effort in analysis and publication of all the data products from this work.

## 7 Broader Impacts

The fundamental goal of the proposed work is to make data easily accessible to individuals who want to answer science questions.

This work is immediately useful to the PI's own collaboration and - if the work is successful - useful to practicing scientists who can more-easily extract information from their datasets.

But this work extends beyond the community of active scientists to scientific learners.

**Technical Development** CDMS and SuperCDMS have developed, and continue to advance detector technologies that have significant impact both inside and outside of the dark matter community. These technologies include electrothermal feedback, transition-edge sensors (TESs); detection and utilization of athermal phonons; large, kilogram-scale phonon-mediated germanium and silicon-based detectors; arrays of large-scale, low-noise SQUIDs; and usage of LPN HEMPTs for ionization readout.

The field of direct detection of dark matter has benefited from these advances. The EDELWEISS and CRESST collaborations have both implemented some of these concepts. In addition, SuperCDMSâĂŹ pioneering of very-low-mass WIMP detection has had a strong impact on the field. There are now many R&D efforts focused on very-low-mass WIMP detection. Multiple of these programs are run by PIs who were primarily engaged in liquid noble direct detection. The concepts developed by SuperCDMS have had far reaching impacts benefiting even direct detection experiments using the liquid nobles.

The broader, basic-science community has benefited from SuperCDMS technology advances. The TES technology is now a core technology for multiple cosmic microwave background experiments including APEX, South Pole Telescope, and PolarBear. It is being rapidly adapted for far-infrared instruments including Super-SCUBA, single-photon spectroscopy, and large X-ray calorimetry experiments at NIST and NASA GSFC including the forthcoming Micro-X, which will launch the first TES into space. In addition, the TES technology is beginning to impact industry as TESs are being adapted for use in electron microscope sensitive surface element analysis and are being investigated for use in quantum coherence and computing.

The direct impact of SuperCDMS research is underscored by the number of graduate students impacted and the number and quality of the SuperCDMS publications. In the past three years, SuperCDMS has produced 12 Ph.D. theses and 11 science publications not including conference proceedings. The impact of these publications is underscored by their many citations (over 3,000 since 2011).

**Undergraduate, Graduate, and Postgraduate Education** Direct detection of dark matter is multidisciplinary and therefore creates a broad spectrum of opportunities for students to explore and advance in many fields. From year to year, numerous undergraduates, graduate students and postdoctoral researchers from the SuperCDMS NSF-funded institutions have benefited from these collaborative opportunities and have been guided to develop precise simulations of semiconductor-based detectors, sensitive analysis techniques, statistical applications, and application of advanced software techniques. These opportunities combined with regular mentoring have afforded many high-visibility talks for students and postdocs.

Sadoulet is the director of Berkeley Connect in Physics, a mentoring program within the physics department that accepts undergraduate students at all levels. The course is a small seminar class led by a physics graduate student, under the guidance of the director. The goals of the program are to help students develop understanding, community, and career preparedness that go beyond what traditional courses provide. Interactions with graduate students and faculty play a large role throughout the semester. Under SadouletâĂŹs leadership, this course has significantly gained in enrollment with every succeeding semester since its launch in Spring 2014.

The Physics Department at CU Denver maintains an undergraduate-only program committed to providing students with hands-on research experience at this urban campus. The Department operates in close collaboration with the Physics Department at Metropolitan State University of Denver, and students from both institutions participate in the CU Denver groupâĂŹs activities, broadening the reach of the research program beyond a single institution. Underrepresented groups and first-generation college students are well represented in the PIâĂŹs group, and many of his mentees have proceeded to graduate school in physics and other disciplines, or embarked on careers in industry at companies such as Northrop Grumman. Currently, there are eight undergraduates in the CU Denver group majoring in Physics, Electrical Engineering, and Mechanical Engineering.

Southern Methodist U. typically funds one to three undergraduate students to work with the Cooley group each year. In the past five years, the eight undergraduates who have participated in the groupâĂŹs research have either entered graduate programs in physics or engineering or embarked on careers in industry at companies such as Texas Instruments and AT&T.

**K-12 Education** SuperCDMS institutions provide strong support for K-12 education through programs for both teachers and students. Too numerous to describe in detail in this proposal, these programs take place in local schools, on campus and directly in the lab. At the core of many of the programs is engagement of students from disadvantaged backgrounds and underrepresented groups, including women.

CU Denver, U. Florida, SMU, and USD engage middle school students in physics learning through lab tours, classroom visits and science camp. Northwestern U., UC Berkeley, U. Florida, TAMU, SMU, and SD School of Mines and Technology support a range of high school programs, including summer research internships for students and teachers, hands-on research projects for underserved high school students of color, mini-courses to prepare Native American students for postsecondary education in science and engineering, interviews with physicists conducted by high school students, and structured physics courses for high school teachers.

## 8 Results from Prior NSF Support

In this section we describe the results from prior NSF support.

**PHY-1809769**
*Collaborative Research: The SuperCDMS SNOLAB Experiment*
Period of Support: 8/2018–7/2021
Amount of Support: $340,000
Publications: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]
Data products from this work are available at the SuperCDMS collaboration's publications website [12].

### 8.0.1 Intellectual merit

This grant supports students and scientists working on an experiment that addresses one of the most fundamental problems of modern science, the nature of dark matter. The SuperCDMS SNOLAB experiment will achieve world-leading sensitivity for dark matter searches in the 1–10 GeV/c$^2$ mass range.

Analyses of CDMS II data demonstrated the power of improved analysis techniques [7, 4], and provided limits for alternate dark matter models [8]. It also revealed a possible signal on Si [13], the analysis and publication of which was lead by the Figueroa Group, then at MIT. Operation of 15 SuperCDMS "iZIP" detectors at Soudan since 2012 demonstrated the detectors' rejection capabilities [14] and yielded world-leading sensitivity to low-mass DM [15, 11, 2].

The first operation of a single ("CDMSlite") detector with a high voltage bias provided the world's most constraining limits for DM masses below 6 GeV/$c^2$ by achieving an extremely low energy threshold of 170 eV electron-recoil energy [15]. A second run of the detector reached an energy threshold for electron recoils as low as 56 eV and demonstrated the power of a fiducialization cut [2]. The Figueroa Group developed key data quality cuts for the first two CDMSlite run analyses, and is currently involved in the analysis of the third run.

### 8.0.2 Accomplishments

The Roberts group is funded primarily for contributions to the data acquisition and data quality systems.

A critical need facing the collaboration as we move to larger data sets is transitioning our analysis platform to computing clusters where we can submit jobs to batch queues. This requires a reworking of existing analysis software, which is primarily MatLab-based and cannot be run at SLAC.

The group began testing and documenting the installation requirements of prototype python software and has since developed the first isolated build environment, allowing reliable installation of the analysis tools across platforms.

Josh Elsarboukh has worked closely with SLAC computing division to successfully deploy this analysis environment via a web interface. This work represents an unprecedented ease of access within the collaboration and has made it possible for test facilities working on crucial R&D and calibration efforts to efficiently analyze their data.

### 8.0.3 Broader impacts

The SuperCDMS experimental and R&D efforts advance phonon-mediated detectors and new active veto concepts, which have already found many applications in cosmology, astronomy and industry.

This grant strongly contributes to the training of undergraduate researchers. The CU Denver Physics Department maintains an undergraduate-only program committed to providing students with hands-on research experience at this urban campus. Electrical Engineering and Mechanical Engineering students have also participated in internships in the CU Denver lab. Minorities and first-generation college students have been well represented in these internships.

## References


[1] R. Agnese, A. J. Anderson, T. Aramaki, I. Arnquist, W. Baker, D. Barker, R. Basu Thakur, D. A. Bauer, A. Borgland, M. A. Bowles, P. L. Brink, R. Bunker, B. Cabrera, D. O. Caldwell, R. Calkins, C. Cartaro, D. G. Cerdeño, H. Chagani, Y. Chen, J. Cooley, B. Cornell, P. Cushman, M. Daal, P. C. F. Di Stefano, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, M. Fritts, G. Gerbier, M. Ghaith, G. L. Godfrey, S. R. Golwala, J. Hall, H. R. Harris, T. Hofer, D. Holmgren, Z. Hong, E. Hoppe, L. Hsu, M. E. Huber, V. Iyer, D. Jardin, A. Jastram, M. H. Kelsey, A. Kennedy, A. Kubik, N. A. Kurinsky, A. Leder, B. Loer, E. Lopez Asamar, P. Lukens, R. Mahapatra, V. Mandic, N. Mast, N. Mirabolfathi, R. A. Moffatt, J. D. Morales Mendoza, J. L. Orrell, S. M. Oser, K. Page, W. A. Page, R. Partridge, M. Pepin, A. Phipps, S. Poudel, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, A. Roberts, A. E. Robinson, H. E. Rogers, T. Saab, B. Sadoulet, J. Sander, K. Schneck, R. W. Schnee, B. Serfass, D. Speller, M. Stein, J. Street, H. A. Tanaka, D. Toback, R. Underwood, A. N. Villano, B. von Krosigk, B. Welliver, J. S. Wilson, D. H. Wright, S. Yellin, J. J. Yen, B. A. Young, X. Zhang, and X. Zhao. Projected Sensitivity of the SuperCDMS SNOLAB experiment. *arXiv:1610.00006 [physics.ins-det]*, September 2016.

[2] SuperCDMS Collaboration, R. Agnese, A. J. Anderson, T. Aramaki, M. Asai, W. Baker, D. Balakishiyeva, D. Barker, R. Basu Thakur, D. A. Bauer, J. Billard, A. Borgland, M. A. Bowles, P. L. Brink, R. Bunker, B. Cabrera, D. O. Caldwell, R. Calkins, D. G. Cerdeno, H. Chagani, Y. Chen, J. Cooley, B. Cornell, P. Cushman, M. Daal, P. C. F. Di Stefano, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, M. Ghaith, G. L. Godfrey, S. R. Golwala, J. Hall, H. R. Harris, T. Hofer, D. Holmgren, L. Hsu, M. E. Huber, D. Jardin, A. Jastram, O. Kamaev, B. Kara, M. H. Kelsey, A. Kennedy, A. Leder, B. Loer, E. Lopez Asamar, P. Lukens, R. Mahapatra, V. Mandic, N. Mast, N. Mirabolfathi, R. A. Moffatt, J. D. Morales Mendoza, S. M. Oser, K. Page, W. A. Page, R. Partridge, M. Pepin, A. Phipps, K. Prasad, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, Y. Ricci, A. Roberts, H. E. Rogers, T. Saab, B. Sadoulet, J. Sander, K. Schneck, R. W. Schnee, S. Scorza, B. Serfass, B. Shank, D. Speller, D. Toback, R. Underwood, S. Upadhyayula, A. N. Villano, B. Welliver, J. S. Wilson, D. H. Wright, S. Yellin, J. J. Yen, B. A. Young, and J. Zhang. New Results from the Search for Low-Mass Weakly Interacting Massive Particles with the CDMS Low Ionization Threshold Experiment. *Phys. Rev. Lett.*, 116(7):071301, 2016.

[3] Ciaran A. J. O'Hare, Anne M. Green, Julien Billard, Enectali Figueroa-Feliciano, and Louis E. Strigari. Readout strategies for directional dark matter detection beyond the neutrino background. *Phys. Rev.*, D92(6):063518, 2015.

[4] R. Agnese, A. J. Anderson, M. Asai, D. Balakishiyeva, D. Barker, R. Basu Thakur, D. A. Bauer, J. Billard, A. Borgland, M. A. Bowles, D. Brandt, P. L. Brink, R. Bunker, B. Cabrera, D. O. Caldwell, R. Calkins, D. G. Cerdeño, H. Chagani, Y. Chen, J. Cooley, B. Cornell, C. H. Crewdson, P. Cushman, M. Daal, P. C. F. Di Stefano, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, G. L. Godfrey, S. R. Golwala, J. Hall, H. R. Harris, S. A. Hertel, T. Hofer, D. Holmgren, L. Hsu, M. E. Huber, D. Jardin, A. Jastram, O. Kamaev, B. Kara, M. H. Kelsey, A. Kennedy, M. Kiveni, K. Koch, A. Leder, B. Loer, E. Lopez Asamar, P. Lukens, R. Mahapatra, V. Mandic, K. A. McCarthy, N. Mirabolfathi, R. A. Moffatt, S. M. Oser, K. Page, W. A. Page, R. Partridge, M. Pepin, A. Phipps, K. Prasad, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, Y. Ricci, H. E. Rogers, T. Saab, B. Sadoulet, J. Sander, K. Schneck, R. W. Schnee, S. Scorza, B. Serfass, B. Shank, D. Speller, D. Toback, S. Upadhyayula, A. N.

Villano, B. Welliver, J. S. Wilson, D. H. Wright, X. Yang, S. Yellin, J. J. Yen, B. A. Young, J. Zhang, and SuperCDMS Collaboration. Improved WIMP-search reach of the CDMS II germanium data. *Physical Review D*, 92(7):072003, October 2015.

[5] K. Schneck, B. Cabrera, D. G. Cerdeño, V. Mandic, H. E. Rogers, R. Agnese, A. J. Anderson, M. Asai, D. Balakishiyeva, D. Barker, R. Basu Thakur, D. A. Bauer, J. Billard, A. Borgland, D. Brandt, P. L. Brink, R. Bunker, D. O. Caldwell, R. Calkins, H. Chagani, Y. Chen, J. Cooley, B. Cornell, C. H. Crewdson, P. Cushman, M. Daal, P. C. F. Di Stefano, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, G. L. Godfrey, S. R. Golwala, J. Hall, H. R. Harris, T. Hofer, D. Holmgren, L. Hsu, M. E. Huber, D. M. Jardin, A. Jastram, O. Kamaev, B. Kara, M. H. Kelsey, A. Kennedy, A. Leder, B. Loer, E. Lopez Asamar, P. Lukens, R. Mahapatra, K. A. McCarthy, N. Mirabolfathi, R. A. Moffatt, J. D. Morales Mendoza, S. M. Oser, K. Page, W. A. Page, R. Partridge, M. Pepin, A. Phipps, K. Prasad, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, Y. Ricci, A. Roberts, T. Saab, B. Sadoulet, J. Sander, R. W. Schnee, S. Scorza, B. Serfass, B. Shank, D. Speller, D. Toback, S. Upadhyayula, A. N. Villano, B. Welliver, J. S. Wilson, D. H. Wright, X. Yang, S. Yellin, J. J. Yen, B. A. Young, J. Zhang, and SuperCDMS Collaboration. Dark matter effective field theory scattering in direct detection experiments. *Physical Review D*, 91(9):092004, May 2015.

[6] M. Pyle, E. Figueroa-Feliciano, and B. Sadoulet. Optimized Designs for Very Low Temperature Massive Calorimeters. *ArXiv e-prints*, March 2015.

[7] R. Agnese, A. J. Anderson, D. Balakishiyeva, R. Basu Thakur, D. A. Bauer, J. Billard, A. Borgland, M. A. Bowles, D. Brandt, P. L. Brink, R. Bunker, B. Cabrera, D. O. Caldwell, D. G. Cerdeno, H. Chagani, Y. Chen, J. Cooley, B. Cornell, C. H. Crewdson, P. Cushman, M. Daal, P. C. F. Di Stefano, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, M. Fritts, G. L. Godfrey, S. R. Golwala, M. Graham, J. Hall, H. R. Harris, S. A. Hertel, T. Hofer, D. Holmgren, L. Hsu, M. E. Huber, A. Jastram, O. Kamaev, B. Kara, M. H. Kelsey, A. Kennedy, M. Kiveni, K. Koch, A. Leder, B. Loer, E. Lopez Asamar, R. Mahapatra, V. Mandic, C. Martinez, K. A. McCarthy, N. Mirabolfathi, R. A. Moffatt, D. C. Moore, R. H. Nelson, S. M. Oser, K. Page, W. A. Page, R. Partridge, M. Pepin, A. Phipps, K. Prasad, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, Y. Ricci, H. E. Rogers, T. Saab, B. Sadoulet, J. Sander, K. Schneck, R. W. Schnee, S. Scorza, B. Serfass, B. Shank, D. Speller, S. Upadhyayula, A. N. Villano, B. Welliver, D. H. Wright, S. Yellin, J. J. Yen, B. A. Young, and J. Zhang. Maximum likelihood analysis of low energy cdms ii germanium data. *Phys. Rev. D*, 91:052021, Mar 2015.

[8] R. Agnese, A. J. Anderson, D. Balakishiyeva, R. Basu Thakur, D. A. Bauer, J. Billard, A. Borgland, M. A. Bowles, D. Brandt, P. L. Brink, R. Bunker, B. Cabrera, D. O. Caldwell, D. G. Cerdeno, H. Chagani, Y. Chen, J. Cooley, B. Cornell, C. H. Crewdson, P. Cushman, M. Daal, P. C. F. Di Stefano, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, G. L. Godfrey, S. R. Golwala, J. Hall, H. R. Harris, S. A. Hertel, T. Hofer, D. Holmgren, L. Hsu, M. E. Huber, A. Jastram, O. Kamaev, B. Kara, M. H. Kelsey, A. Kennedy, M. Kiveni, K. Koch, A. Leder, B. Loer, E. Lopez Asamar, R. Mahapatra, V. Mandic, C. Martinez, K. A. McCarthy, N. Mirabolfathi, R. A. Moffatt, D. C. Moore, H. Nelson, R. H. Nelson, R. W. Ogburn, K. Page, W. A. Page, R. Partridge, M. Pepin, A. Phipps, K. Prasad, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, Y. Ricci, H. E. Rogers, T. Saab, B. Sadoulet, J. Sander, K. Schneck, R. W. Schnee, S. Scorza, B. Serfass, B. Shank, D. Speller, S. Upadhyayula, A. N. Villano, B. Welliver, D. H. Wright, S. Yellin, J. J. Yen, B. A. Young, J. Zhang, and CDMS Collabora-

tion. First direct limits on Lightly Ionizing Particles with electric charge less than $e/6$. *Phys. Rev. Lett.*, 114(11):111302, 2015.

[9] J. Billard, L. E. Strigari, and E. Figueroa-Feliciano. Solar neutrino physics with low-threshold dark matter detectors. *Phys. Rev.*, D91(9):095023, May 2015.

[10] F. Ruppin, J. Billard, E. Figueroa-Feliciano, and L. Strigari. Complementarity of dark matter detectors in light of the neutrino background. *Phys. Rev.*, D90(8):083510, 2014.

[11] R. Agnese, A. J. Anderson, M. Asai, D. Balakishiyeva, R. Basu Thakur, D. A. Bauer, J. Beaty, J. Billard, A. Borgland, M. A. Bowles, D. Brandt, P. L. Brink, R. Bunker, B. Cabrera, D. O. Caldwell, D. G. Cerdeno, H. Chagani, Y. Chen, M. Cherry, J. Cooley, B. Cornell, C. H. Crewdson, P. Cushman, M. Daal, D. DeVaney, P. C. F. Di Stefano, E. Do Couto E Silva, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, G. L. Godfrey, S. R. Golwala, J. Hall, S. Hansen, H. R. Harris, S. A. Hertel, B. A. Hines, T. Hofer, D. Holmgren, L. Hsu, M. E. Huber, A. Jastram, O. Kamaev, B. Kara, M. H. Kelsey, S. Kenany, A. Kennedy, M. Kiveni, K. Koch, A. Leder, B. Loer, E. Lopez Asamar, R. Mahapatra, V. Mandic, C. Martinez, K. A. McCarthy, N. Mirabolfathi, R. A. Moffatt, R. H. Nelson, L. Novak, K. Page, R. Partridge, M. Pepin, A. Phipps, M. Platt, K. Prasad, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, R. W. Resch, Y. Ricci, M. Ruschman, T. Saab, B. Sadoulet, J. Sander, R. L. Schmitt, K. Schneck, R. W. Schnee, S. Scorza, D. N. Seitz, B. Serfass, B. Shank, D. Speller, A. Tomada, S. Upadhyayula, A. N. Villano, B. Welliver, D. H. Wright, S. Yellin, J. J. Yen, B. A. Young, and J. Zhang. Search for low-mass weakly interacting massive particles with supercdms. *Phys. Rev. Lett.*, 112:241302, Jun 2014.

[12] Cdms publications: http://cdms.berkeley.edu/publications.html.

[13] R Agnese, Z Ahmed, A J Anderson, S Arrenberg, D Balakishiyeva, R Basu Thakur, D A Bauer, J Billard, A Borgland, D Brandt, P L Brink, T Bruch, R Bunker, B Cabrera, D O Caldwell, D G Cerdeno, H Chagani, J Cooley, B Cornell, C H Crewdson, P Cushman, M Daal, F Dejongh, E do Couto e Silva, T Doughty, L Esteban, S Fallows, E Figueroa-Feliciano, J Filippini, J Fox, M Fritts, G L Godfrey, S R Golwala, J Hall, R H Harris, S A Hertel, T Hofer, D Holmgren, L Hsu, M E Huber, A Jastram, O Kamaev, B Kara, M H Kelsey, A Kennedy, P Kim, M Kiveni, K Koch, M Kos, S W Leman, B Loer, E Lopez Asamar, R Mahapatra, V Mandic, C Martinez, K A McCarthy, N Mirabolfathi, R A Moffatt, D C Moore, P Nadeau, R H Nelson, K Page, R Partridge, M Pepin, A Phipps, K Prasad, M Pyle, H Qiu, W Rau, P Redl, A Reisetter, Y Ricci, T Saab, B Sadoulet, J Sander, K Schneck, R W Schnee, S Scorza, B Serfass, B Shank, D Speller, K M Sundqvist, A N Villano, B Welliver, D H Wright, S Yellin, J J Yen, J Yoo, B A Young, and J Zhang. Silicon Detector Dark Matter Results from the Final Exposure of CDMS II. *Physical Review Letters*, 111(2):251301, December 2013.

[14] R Agnese, A J Anderson, D Balakishiyeva, R Basu Thakur, D A Bauer, A Borgland, D Brandt, P L Brink, R Bunker, B Cabrera, D O Caldwell, D G Cerdeno, H Chagani, M Cherry, J Cooley, B Cornell, C H Crewdson, P Cushman, M Daal, P C F Di Stefano, E do Couto e Silva, T Doughty, L Esteban, S Fallows, E Figueroa-Feliciano, J Fox, M Fritts, G L Godfrey, S R Golwala, J Hall, H R Harris, J Hasi, S A Hertel, B A Hines, T Hofer, D Holmgren, L Hsu, M E Huber, A Jastram, O Kamaev, B Kara, M H Kelsey, S A Kenany, A Kennedy, C J Kenney, M Kiveni, K Koch, B Loer, E Lopez Asamar, R Mahapatra, V Mandic, C Martinez, K A McCarthy, N Mirabolfathi, R A Moffatt, D C Moore, P Nadeau, R H Nelson, L Novak, K Page, R Partridge, M Pepin, A Phipps, K Prasad, M Pyle, H Qiu, R Radpour, W Rau, P Redl,

A Reisetter, R W Resch, Y Ricci, T Saab, B Sadoulet, J Sander, R Schmitt, K Schneck, R W Schnee, S Scorza, D Seitz, B Serfass, B Shank, D Speller, A Tomada, A N Villano, B Welliver, D H Wright, S Yellin, J J Yen, B A Young, J Zhang, and The SuperCDMS Collaboration. Demonstration of surface electron rejection with interleaved germanium detectors for dark matter searches. *Appl. Phys. Lett.*, 103(16):164105, 2013.

[15] R Agnese, A J Anderson, M Asai, D Balakishiyeva, R Basu Thakur, D A Bauer, J Billard, A Borgland, M A Bowles, D Brandt, P L Brink, R Bunker, B Cabrera, D O Caldwell, D G Cerdeno, H Chagani, J Cooley, B Cornell, C H Crewdson, P Cushman, M Daal, P C F Di Stefano, T Doughty, L Esteban, S Fallows, E Figueroa-Feliciano, G L Godfrey, S R Golwala, J Hall, H R Harris, S A Hertel, T Hofer, D Holmgren, L Hsu, M E Huber, A Jastram, O Kamaev, B Kara, M H Kelsey, A Kennedy, M Kiveni, K Koch, B Loer, E Lopez Asamar, R Mahapatra, V Mandic, C Martinez, K A McCarthy, N Mirabolfathi, R A Moffatt, D C Moore, P Nadeau, R H Nelson, K Page, R Partridge, M Pepin, A Phipps, K Prasad, M Pyle, H Qiu, W Rau, P Redl, A Reisetter, Y Ricci, T Saab, B Sadoulet, J Sander, K Schneck, R W Schnee, S Scorza, B Serfass, B Shank, D Speller, A N Villano, B Welliver, D H Wright, S Yellin, J J Yen, B A Young, and J Zhang. CDMSlite: A Search for Low-Mass WIMPs using Voltage-Assisted Calorimetric Ionization Detection in the SuperCDMS Experiment. *Physical Review Letters*, 112(4):041302, September 2013.