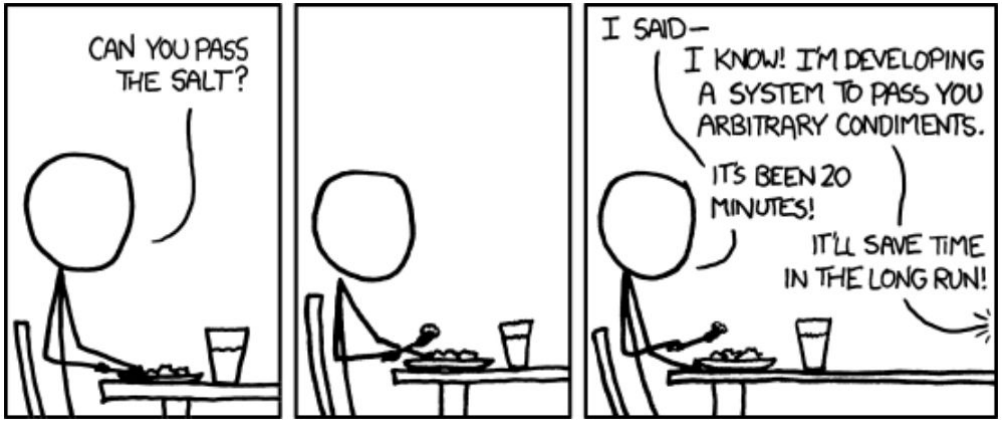
A Dark Matter Perspective

On community software efforts

whoami?

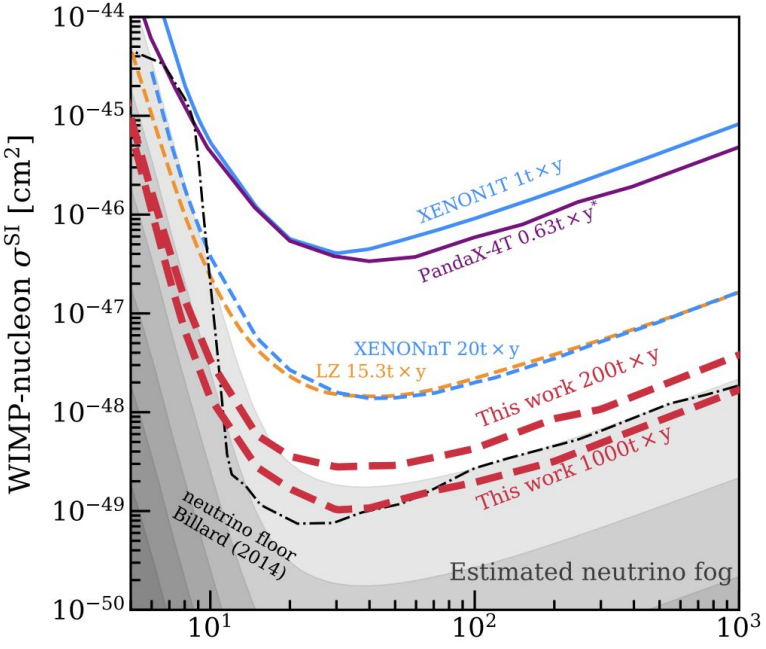
- PI with the SuperCDMS collaboration, assistant professor at CU Denver
- Cryogenic Dark Matter Search
- I'm a huge fan of collaboration and have to remind myself that it does take extra time and that the motivation needs to be solid!



https://xkcd.com/974

Dark Matter Physics Background

- We don't know what our signal will look like. Different experiments focus on different mass ranges.
- Most of these are exclusion plots. Convincing evidence of dark matter discovery will require exquisite understanding of backgrounds and detector response.



From 2203.02309.pdf (arxiv.org)

Physics -> Computing Challenges

- **We don't know what our signal will look like.**
  - Maximizing sensitivity over as large a mass range as possible is a common experimental goal
  - **Many dark matter experiments are pushing hard against their noise wall**
  - Where there's a higher trigger rate
  - And strong incentive to get information on low-frequency noise by storing raw pulses that are as long as possible. Six seconds of digitized pulse data is a serious dream within my collaboration!
- **Convincing evidence of dark matter discovery will require exquisite understanding of backgrounds and detector response.**
  - **Simulations are a critical part of discovery for dark matter searches.** We need experimental and software R&D, together, to be ready for new data.
  - Calibration and simulation are a huge part of analysis and these needs vary across dark matter experiments based on their chosen signals and materials. Despite reasonably intense efforts to record metadata and versioning, **we struggle to reproduce analyses**.
  - Low-energy interactions, at energies approaching the eV - 100 eV scale, are not well studied for many materials and the simulations are correspondingly untrustworthy.
  - We often try neutrons to calibrate the nuclear-recoil scale and getting accurate neutron simulation data is tough. GEANT wasn't built for low-energy neutron accuracy!

Dark Matter: How is it different from HEP?

- No centralization of software support
- No centralized CPU support, disk support, services support like metadata and wikis
  - Some dark matter experiments do fine with data, processing, and analysis all located at a single site. This works best when the single site is not in the US or is at a university.
  - Some experiments need multiple sites to get all the processing and analyzing resources they need
- No centralized support for archival efforts
- Smaller datasets, but different types of analysis (?)
  - My understanding of HEP analyses: most analyzers work with small data sets that have been downselected.
  - Dark matter analysis often require looking at entire, processed datasets to determine the downselection. We have a strong need for tools that can handle large data sets in memory-efficient ways
  - We often return to our lowest-level data for reprocessing

Dark Matter: How is it similar to HEP?

- **I would recommend assuming that if it's a need in HEP, it's a need in Dark Matter**
- So this is not a complete list!
- Training needs - especially for foundation skills - overlap
- We struggle with discoverability and coordinating efforts - less of an issue in HEP!
- Data catalogs are always helpful and sometimes necessary
- Need to process and track data
- Struggle with reproducibility and adequate metadata for calibration, simulation, data processing, and data analysis
- Processed data consists of a list of physics events that each hold physics information. Some of us can manage using traditional arrays.
- For some analyses (machine learning using raw signals), we can never load full datasets into memory. For other analyses, we can load full datasets into memory only once we're at the parameter-extraction stage, sometimes a little earlier. This is a change from the past, when patience and large-RAM machines were a reasonable analysis method.

Training

- DANCE-Edu is an effort from three dark-matter collaborations to combine training effort; ran a workshop together with IRIS-HEP at SNOWMASS
  - DANCE/CoDaS@Snowmass 2022 computational and data science software training (19-July 23, 2022): Overview · Indico (cern.ch)
  - Chris Tunnel, Amy Roberts, Andrew Renshaw, also DANCEOrg on slack
- Open questions about training
  - Are there ways to make training more sustainable and effective?
    - Computational biology has QUBES - Home (qubeshub.org)
    - The team behind QUBES provides contributors with material useful for career advancement
    - When people submit material, the site gives them a way to show use
  - Are there low-impact ways to assess training? Do we need education specialists on board?
  - How do people learn about training opportunities?
    - A problem in the fragmented field of dark matter
    - And there's a lot of training that's useful for dark matter scientists!
      - Cluster-run training for HPC use, profiling, parallel programming, and other "advanced" topics. I only hear about these through the Campus Champions mailing list.
      - HEP, Nuclear Physics (e.g. BAND Framework - Framework) both have significant overlap, BSSw aims to be a resource for scientific software best practices

Making tools more widely usable I

- Software needs overlap in other disciplines - nuclear physics uses uproot heavily, pangeo is addressing many of the problems we need to address in dark matter, astrophysics has similar metadata issues ...
- PONDD is a project that aims to make tools like uproot usable for data with an arbitrary binary format
  - A human describes the data and then the Kaitai compiler generates code that reads this data.
  - We want researchers to write less custom code to analyze their custom data and instead be able to use standard tooling once they describe their data
  - The plan is to use XSEDE/ACCESS resources to spin up the infrastructure needed for the system.
  - Contacts: Amy Roberts, Matt Turk, #midscale-data-access on IRIS-HEP slack
  - It's been tough to get buy-in! We've had to go shopping for people who are stuck enough to be motivated to participate and most of these people are not in dark matter!
- When does it make sense to expand our efforts beyond the Cosmic Frontier?

Making tools more widely usable II

- I think the dark matter community has software overlap with HEP in ways that can have community buy-in
- An example: SuperCDMS is moving to containerization for distributing software builds and this has been amazing for our developers. All of the clusters we work on support Singularity. We've used these containers as the base for JupyterLab instances for analyzers and now people make their first plots in weeks rather than months. The collaboration grouchies have reluctantly admitted that the system is an improvement.
- The dark matter field is about to start taking an order of magnitude more data and many of us need help - most of us have very small computing teams (3 - 4 people who are not dedicated to software) and this hasn't expanded with our data volume.
- There is interest in the dark matter community in coordinated efforts - 26 people contributed to the "Common software needs in dark matter" whitepaper.
- **How can we best work together? How does funding for these projects work in HEP? Could each project grant for dark matter have 20% for cyberinfrastructure built in and paid by OAC? Could OAC have supplements for smaller experiments to support integration with existing ecosystems?**

Thank you!

- Questions, comments?