Trusting dark matter results

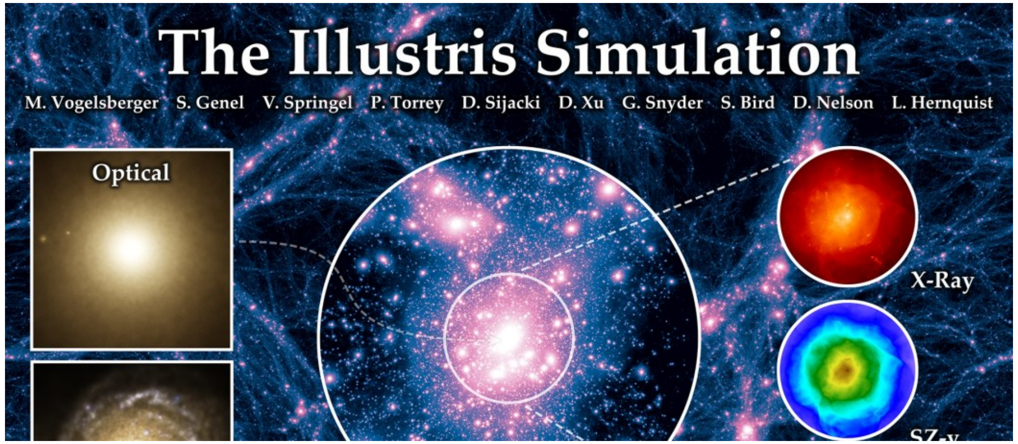
THREE SHORT STORIES ABOUT INCREASING TRUST IN THE DIFFICULT REALM OF DARK MATTER ANALYSIS

CONTACT: AMY.ROBERTS@UCDENVER.EDU

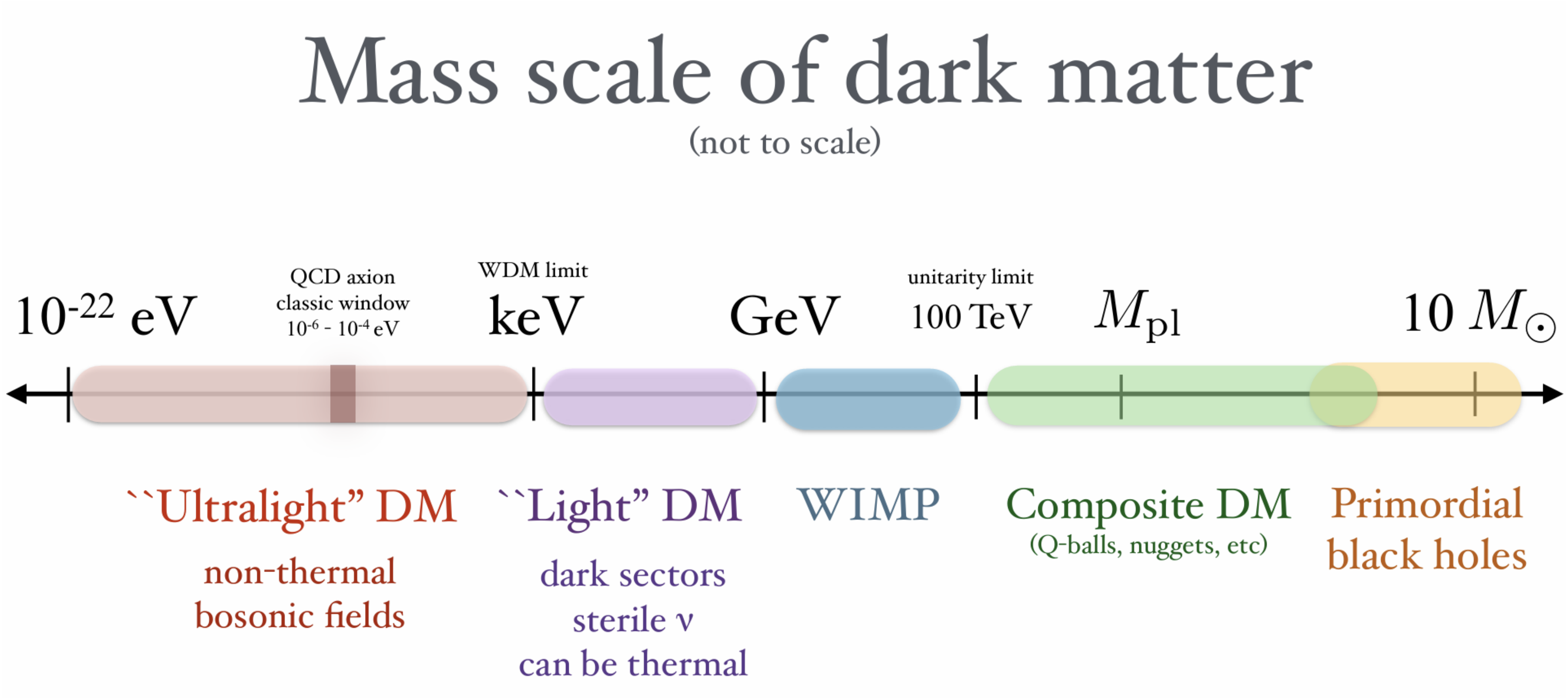
Dark Matter

A simulated universe with a weakly-interacting, ~~40 GeV particle~~ ~0.5 MeV/c^2 mass density looks a lot like our universe.

http://www.illustris-project.org/movies/illustris_movie_rot_sub_frame.mp4



Dark Matter

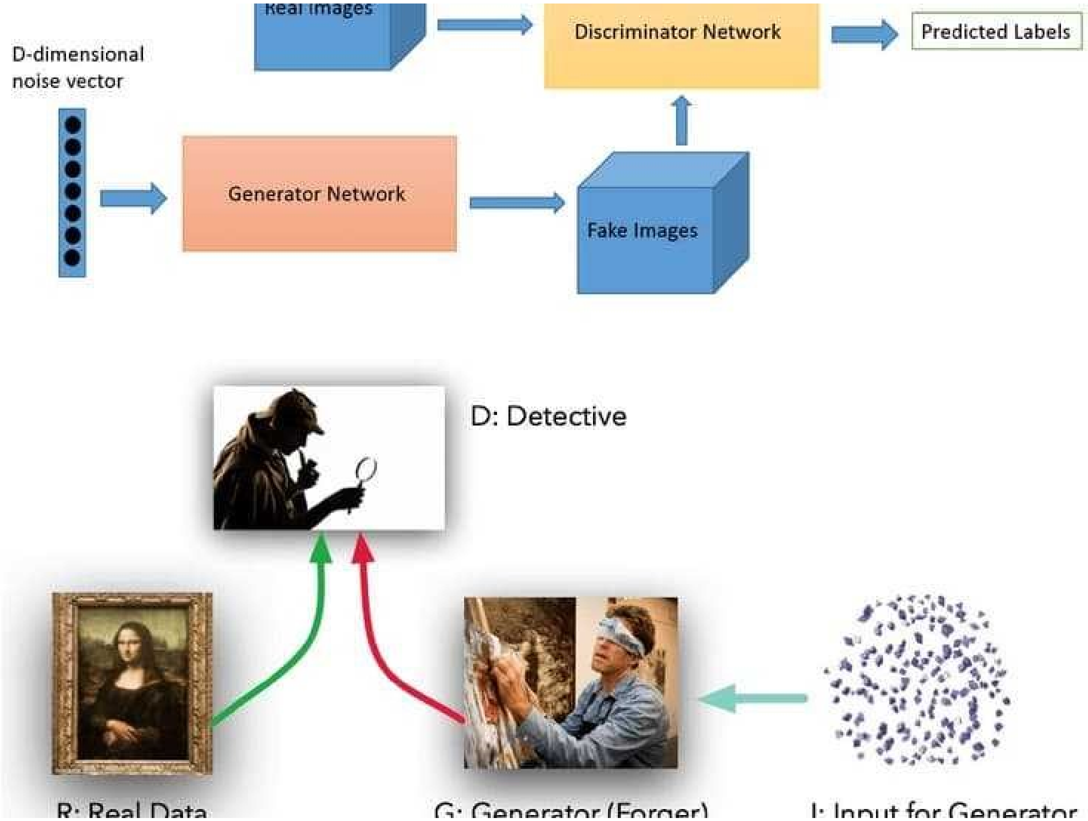


There are many dark matter candidates: fun for detector designing, difficult to blind analysis
Image from [1904.07915 (arxiv.org)]

Story 1. How can we mitigate bias in our analyses?

- We don't know what to expect for a dark matter signal. We do know the models rarely involve peaks :(
- Several dark matter collaborations (SuperCDMS, LUX) have used "salting" to blind their data
- See Maria-Elena Monzani's excellent talk at PHYSTAT-2019-Monzani.pdf (cern.ch) for additional details
- Create "fake" signal data and insert it into the data stream. You should recover the injected signal when you analyze. This is a way to build trust in the analysis.
- This is great! We can create salt for all our signal models!
- **The problem** is that creating salt takes a long time

Story 1. How can we mitigate bias in our analyses?



Generative Adversarial Networks (GANs) are specifically built to create fake data.

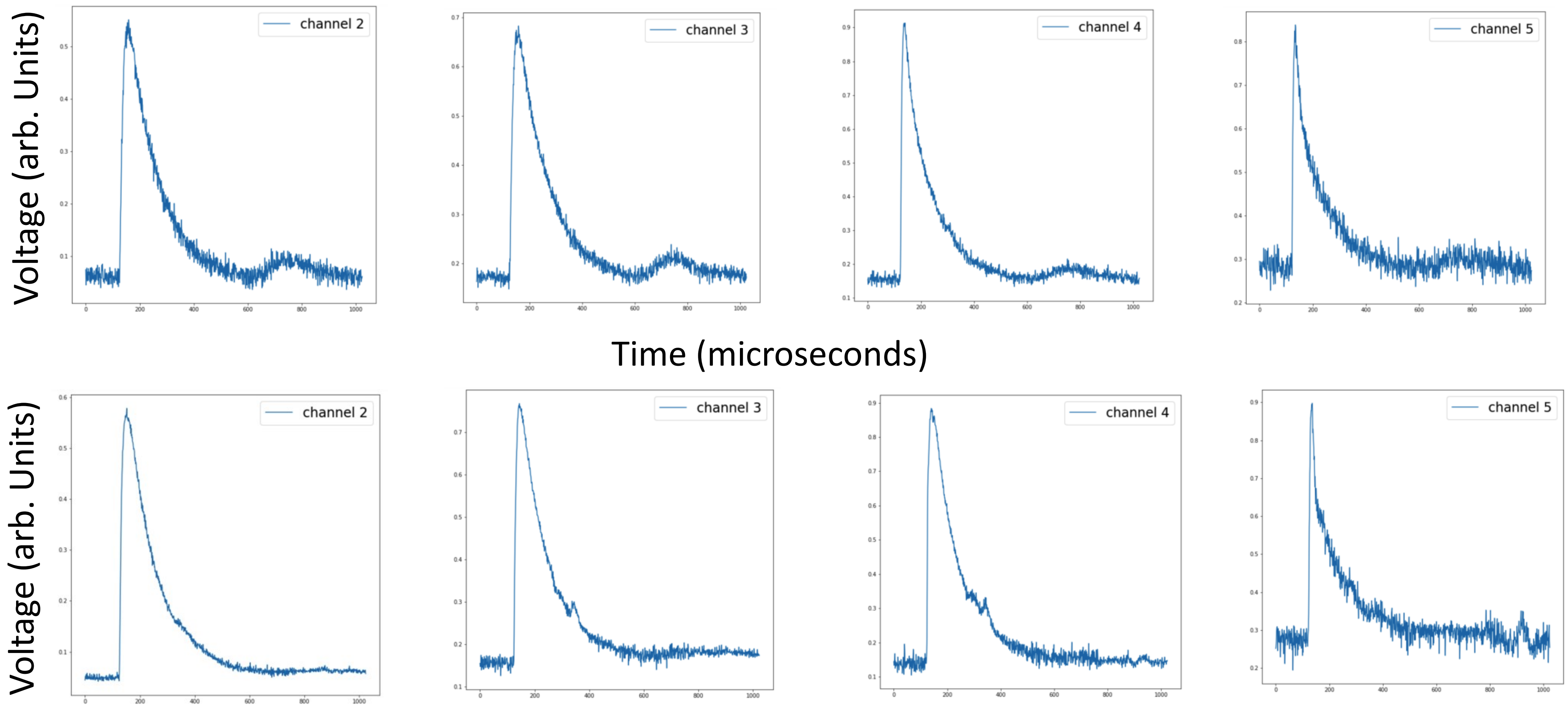
We're working on using GANs to create raw data that's useful as salt.

From Decrypt Generative Adversarial Networks (GAN) | AI Summer (theaisummer.com), with elements from Generative Adversarial Networks for beginners – O'Reilly (oreilly.com)

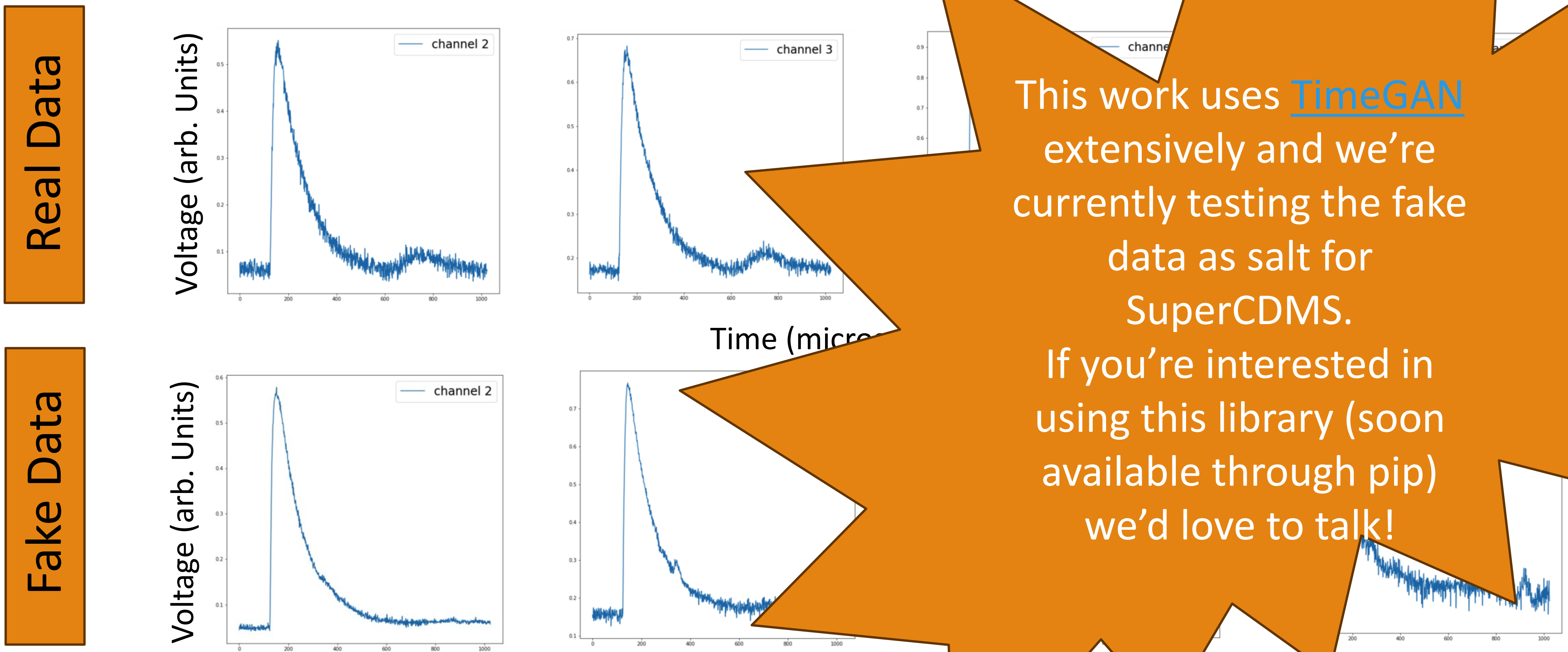
Story 1. How can we mitigate bias in our analyses?

Real Data

Fake Data



Story 1. How can we mitigate bias in our analyses?

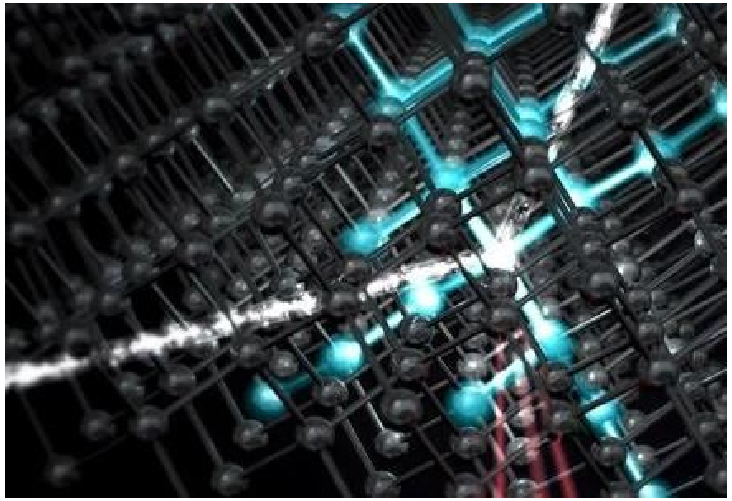


This work uses TimeGAN extensively and we're currently testing the fake data as salt for SuperCDMS.
If you're interested in using this library (soon available through pip) we'd love to talk!

Story 2. What are our detectors telling us?

SUBTEXT: DATA AND ANALYSIS PRESERVATION CAN HAVE SIGNIFICANT SCIENTIFIC VALUE.

WE ARE USING DATA MORE THAN 10 YEARS OLD TO UNDERSTAND GERMANIUM RESPONSE TO DEPOSITED ENERGY!



A particle deposits energy by interacting with a nucleus, and that energy turns into vibration (phonons) and charge (ionization). To reconstruct the energy, we need the expected ratio (yield). The fraction that goes into phonons itself has a variance, labeled "F" in the figure.

Imagine we see an event at 3.5 e/h pairs. How do we interpret this?
With increasingly high resolutions, understanding energy response is increasingly important.

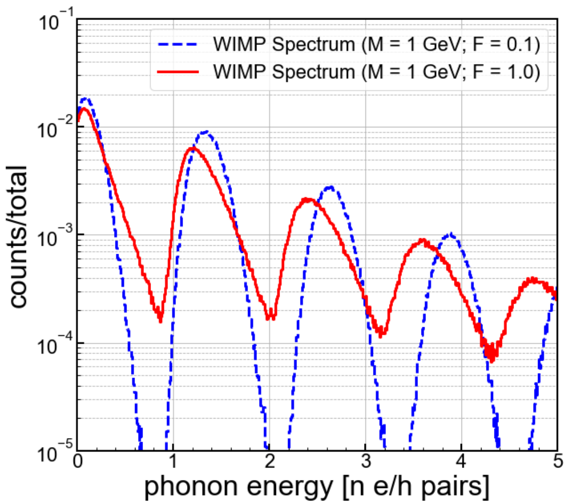


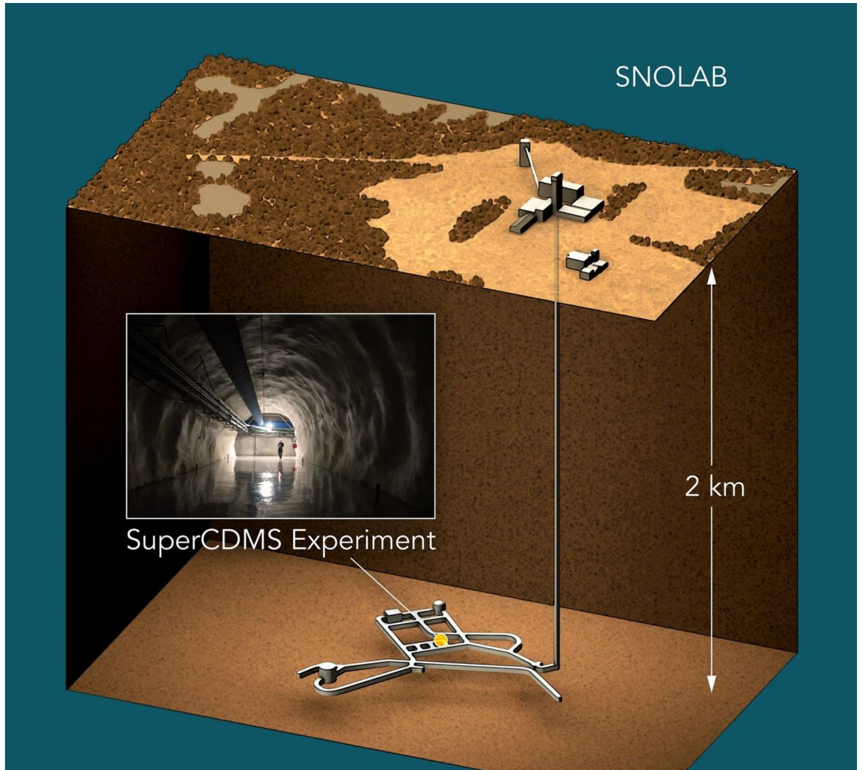
Figure from Anthony Villano

Story 3. How can we constrain our backgrounds?

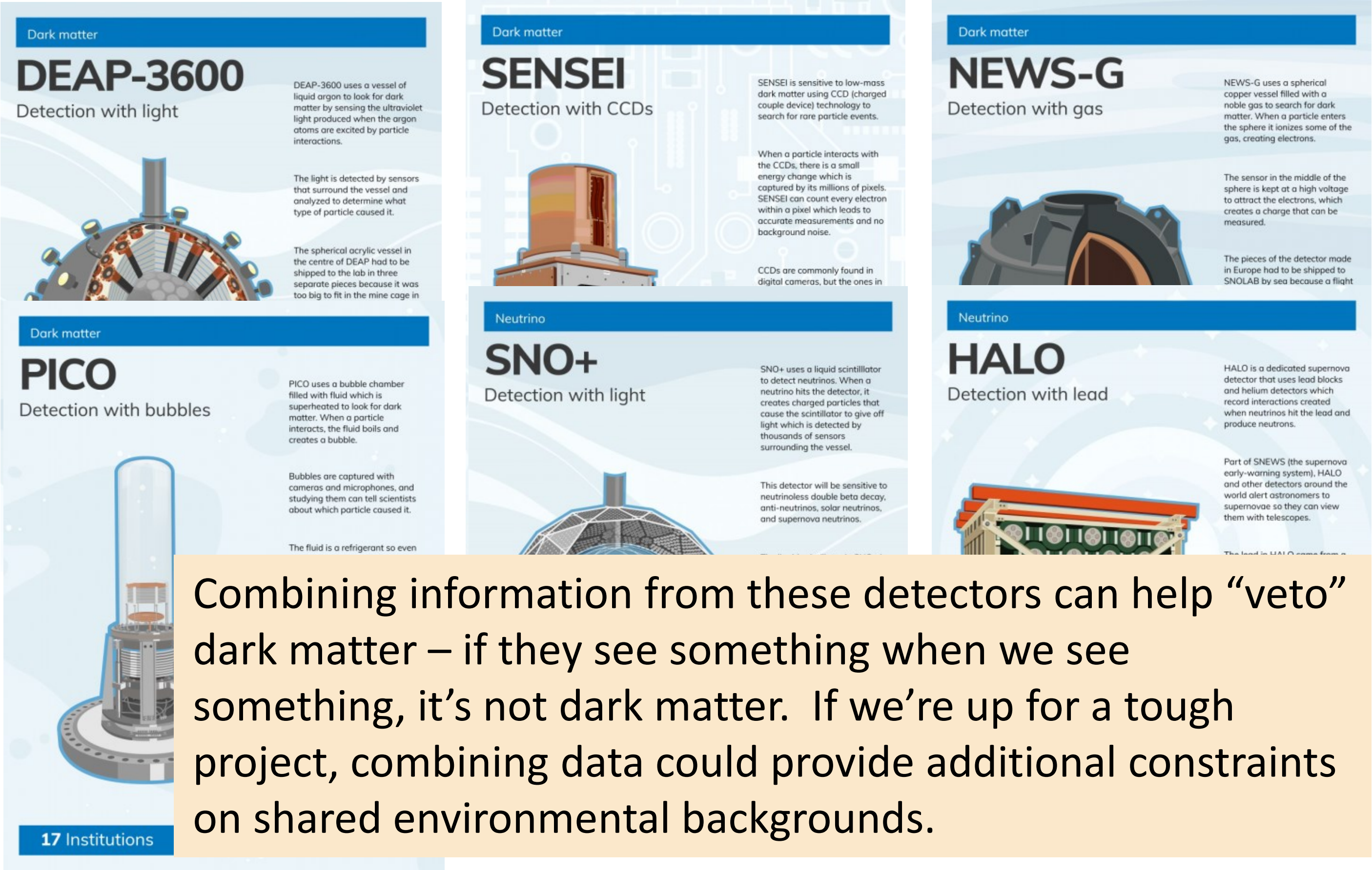
Some of our backgrounds are specific to our experiment.

But other backgrounds are shared by everyone at an underground facility. If I see a germanium nucleus recoil, is it a neutron or something more interesting?

Luckily, there are many detectors near our dark matter detector: [SNOLAB experiment posters | SNOLAB]



[The Search for Dark Matter Continues, More Than a Mile Underground (popularmechanics.com)]

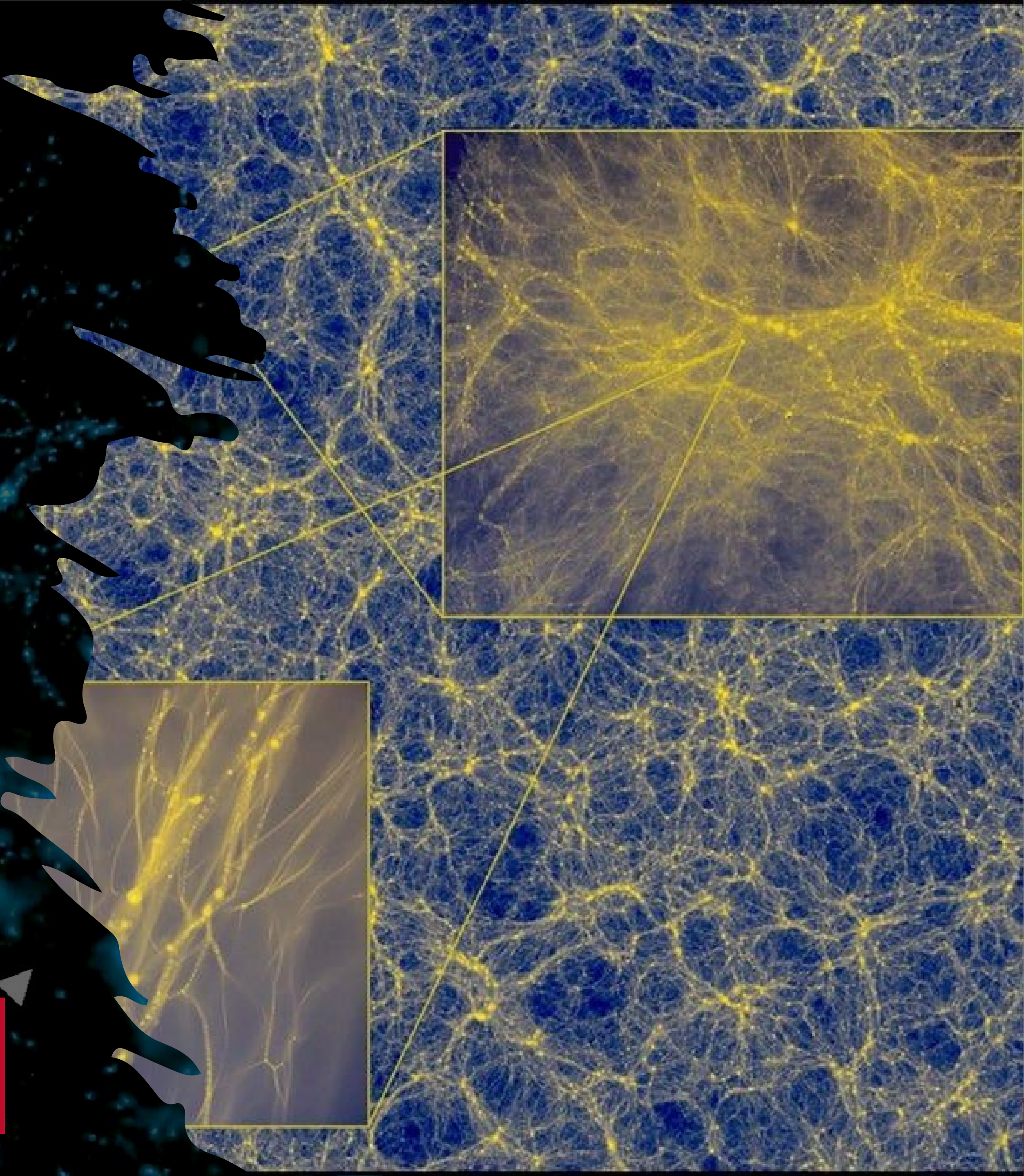


Combining information from these detectors can help “veto” dark matter – if they see something when we see something, it’s not dark matter. If we’re up for a tough project, combining data could provide additional constraints on shared environmental backgrounds.

NSDF-SLAC

Stanford Linear Accelerator Center (**SLAC**)
Super Cryogenic Dark Matter Search (**SuperCDMS**)
Sudbury Neutrino Observatory Lab (**SNOLAB**)
National Science Data Fabric (**NSDF**)

04 MAR 2024

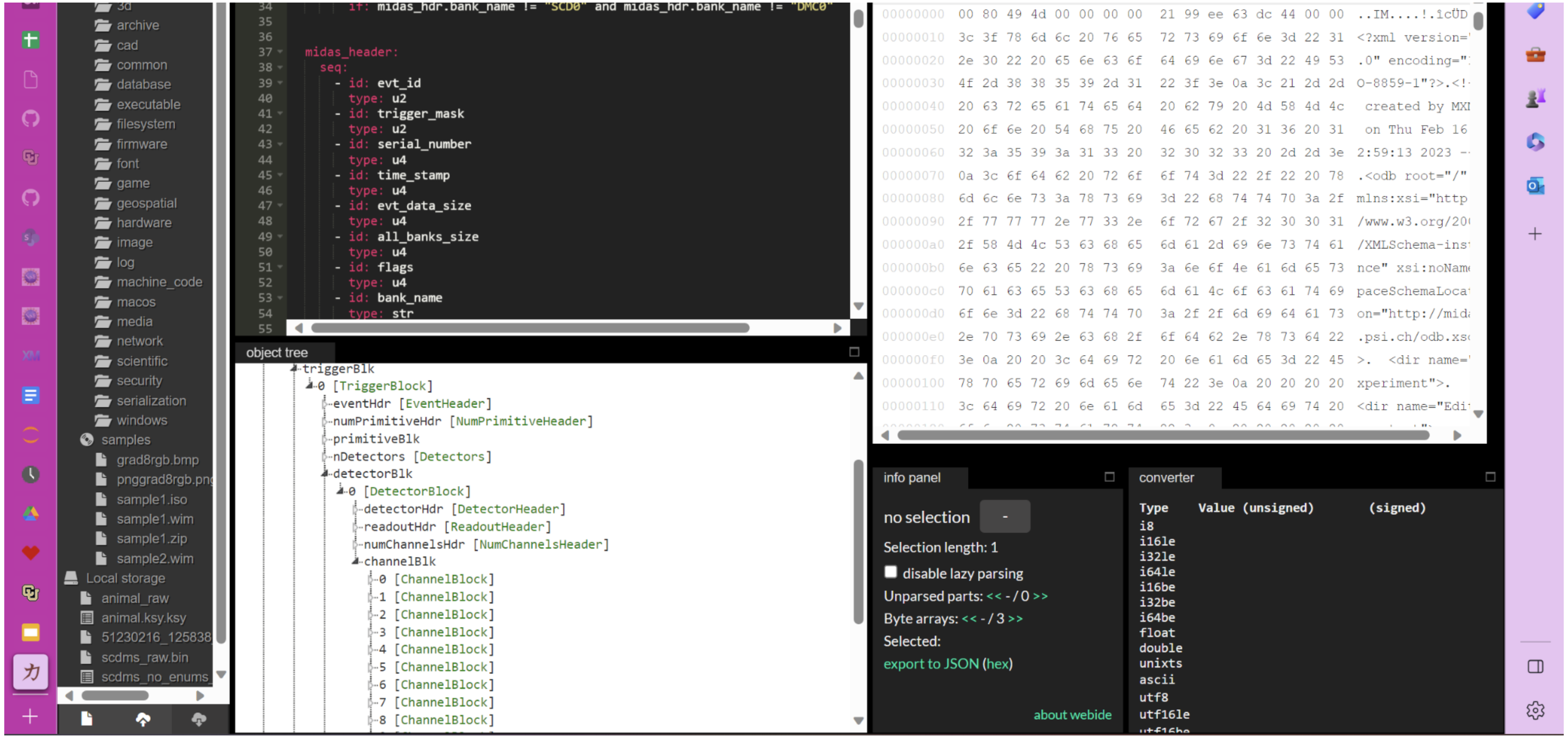


NSDF

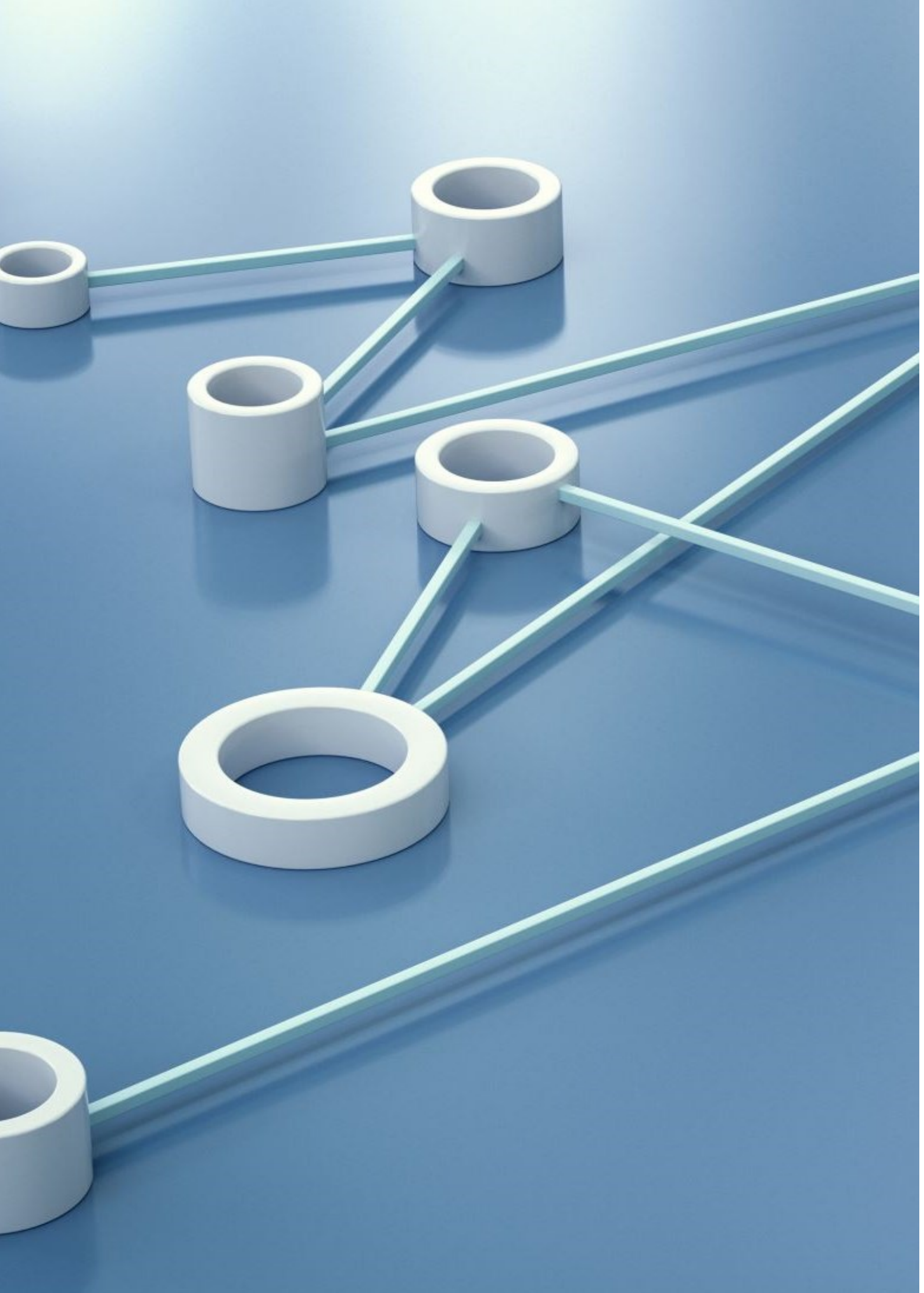
SLAC
NATIONAL ACCELERATOR LABORATORY

: 1.2 billion years

Describing data formats to get a common interface



The PONDD project is a collaborative project to improve access to custom binary data sets. Support for small, non-standard, binary datasets already exists with open-source projects like Kaitai and DFDL. The PONDD project aims to add similar support for GB-scale files.



In conclusion

- Preserving data can lead to better science, particularly as we develop new detectors and analysis methods. Also we need to train our students!
- Sharing data across collaborations is becoming technically feasible and offers new ways to constrain backgrounds.
- Please contact me if you're interested in collaborating!