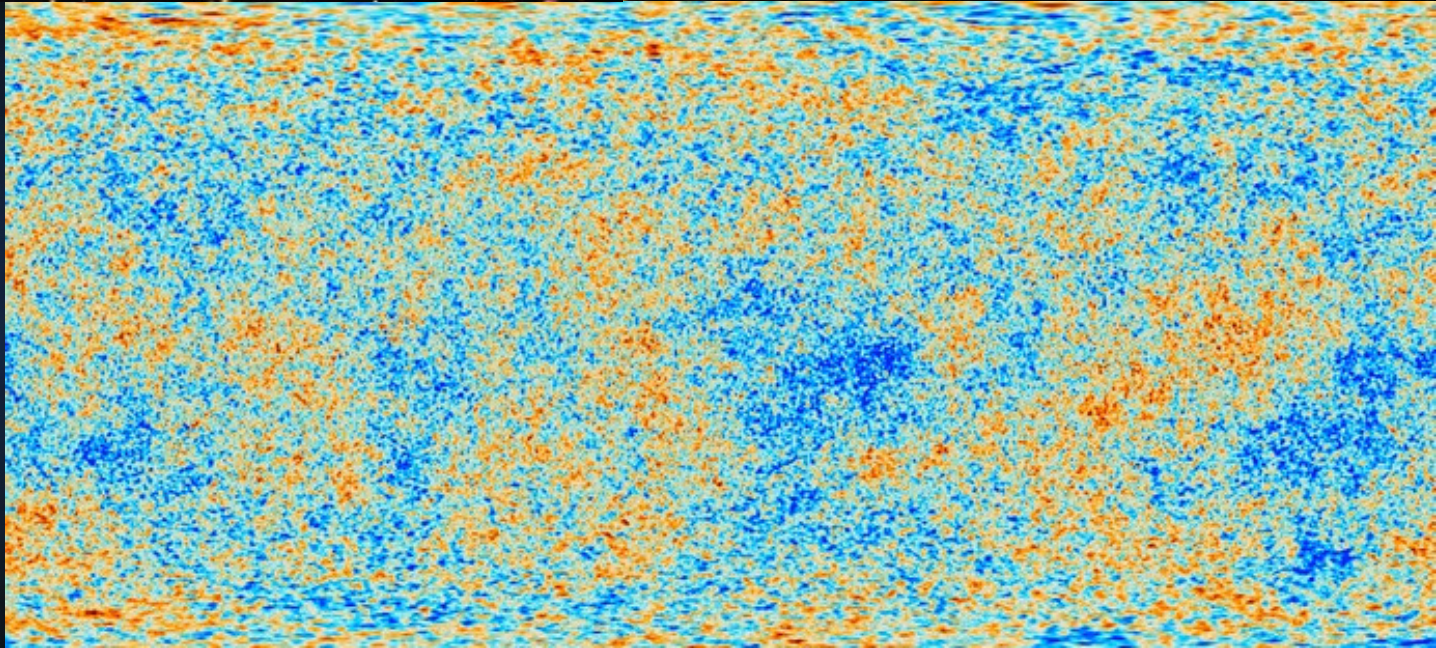
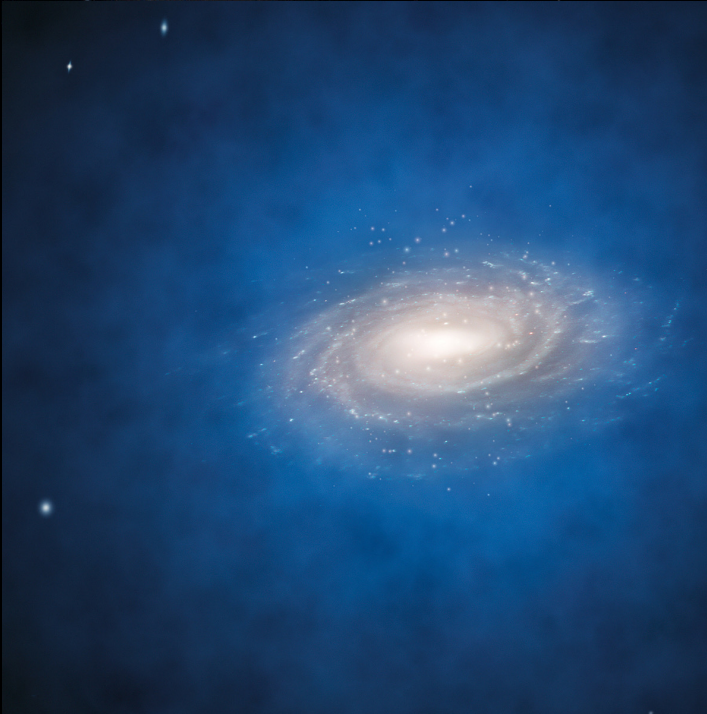
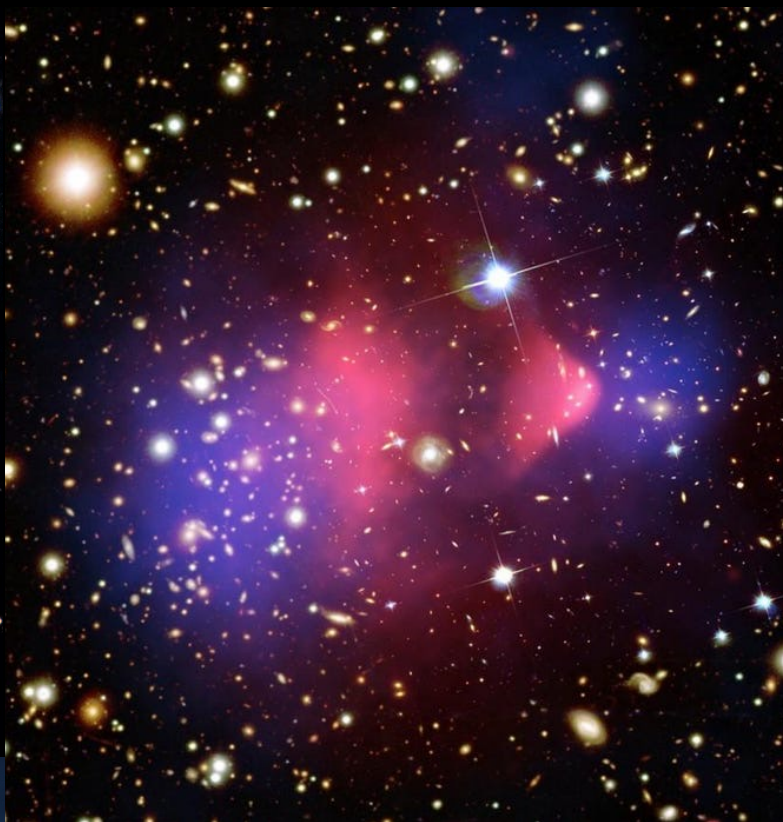


DATA INTENSIVE SEARCHES FOR DARK MATTER WITH LUX-ZEPLIN (LZ)

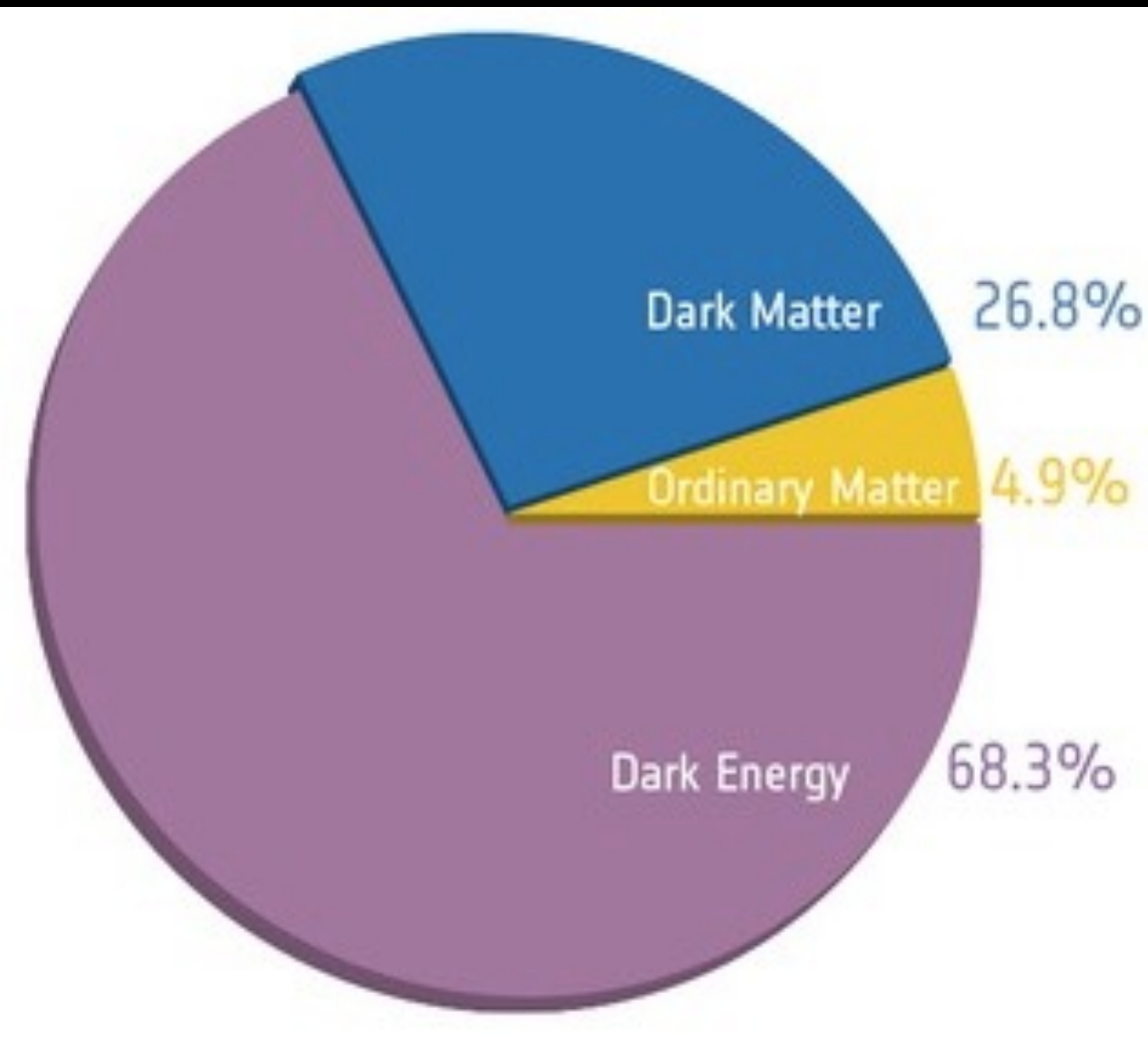


**MARIA ELENA MONZANI, SLAC/STANFORD
CHEP, NORFOLK VA, MAY 11 2023**

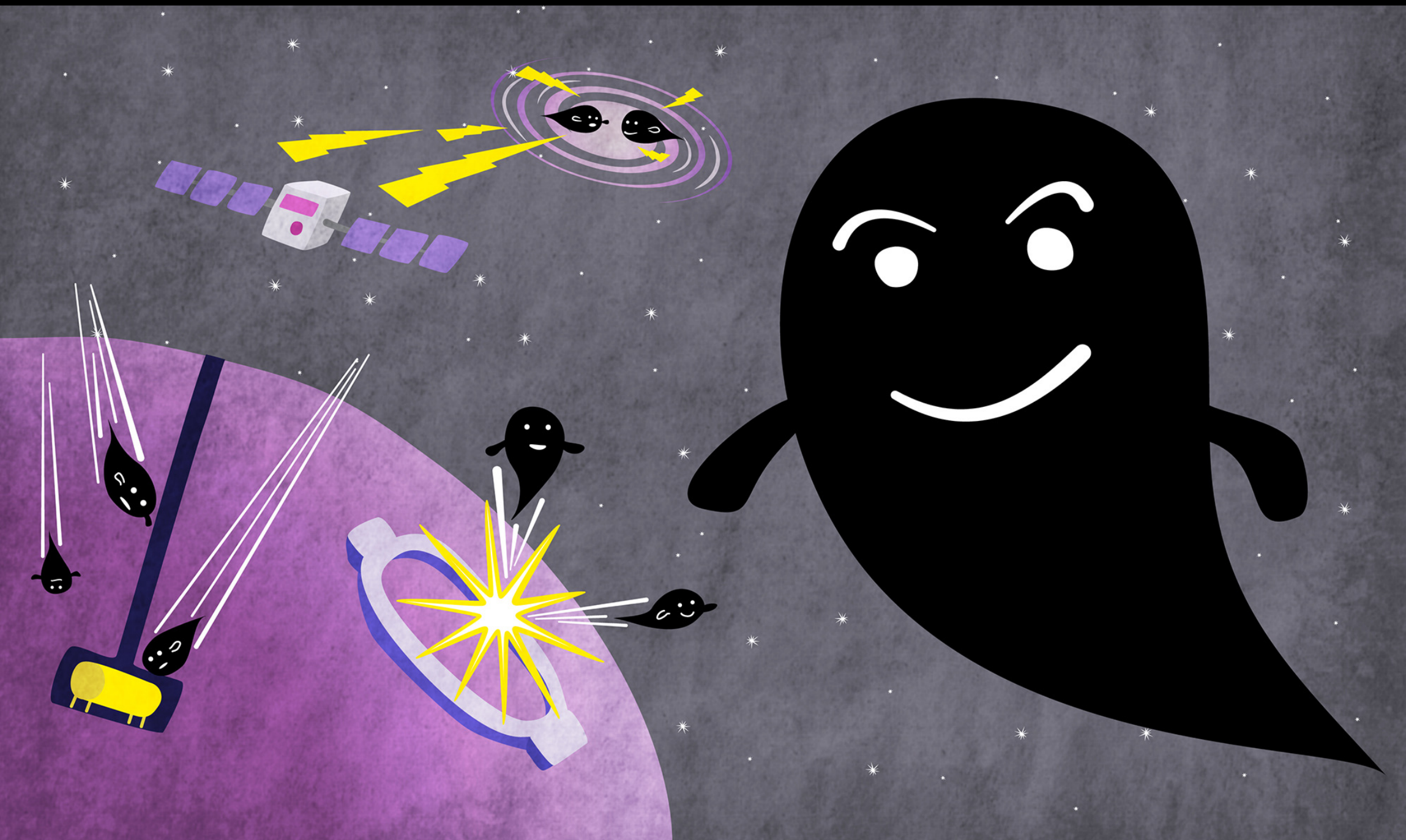
DARK MATTER IS ALL OVER THE PLACE!



LOTS OF WEIRD STUFF OUT THERE!

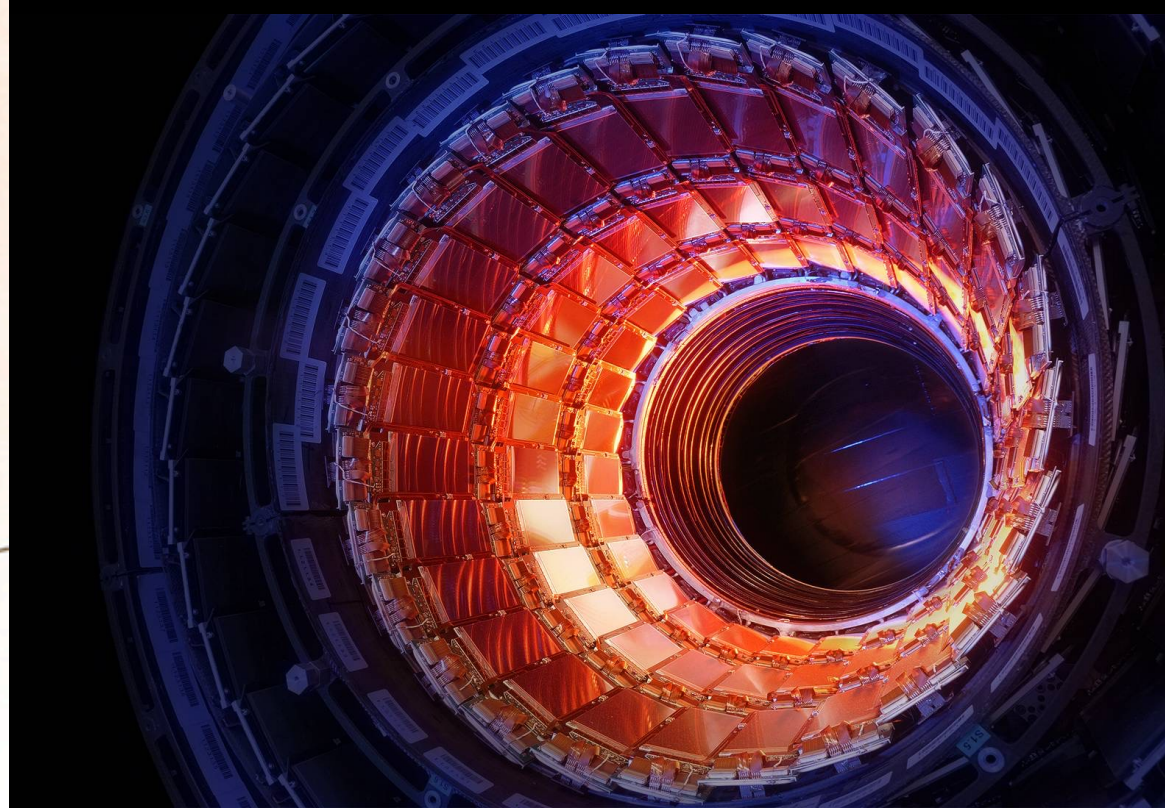
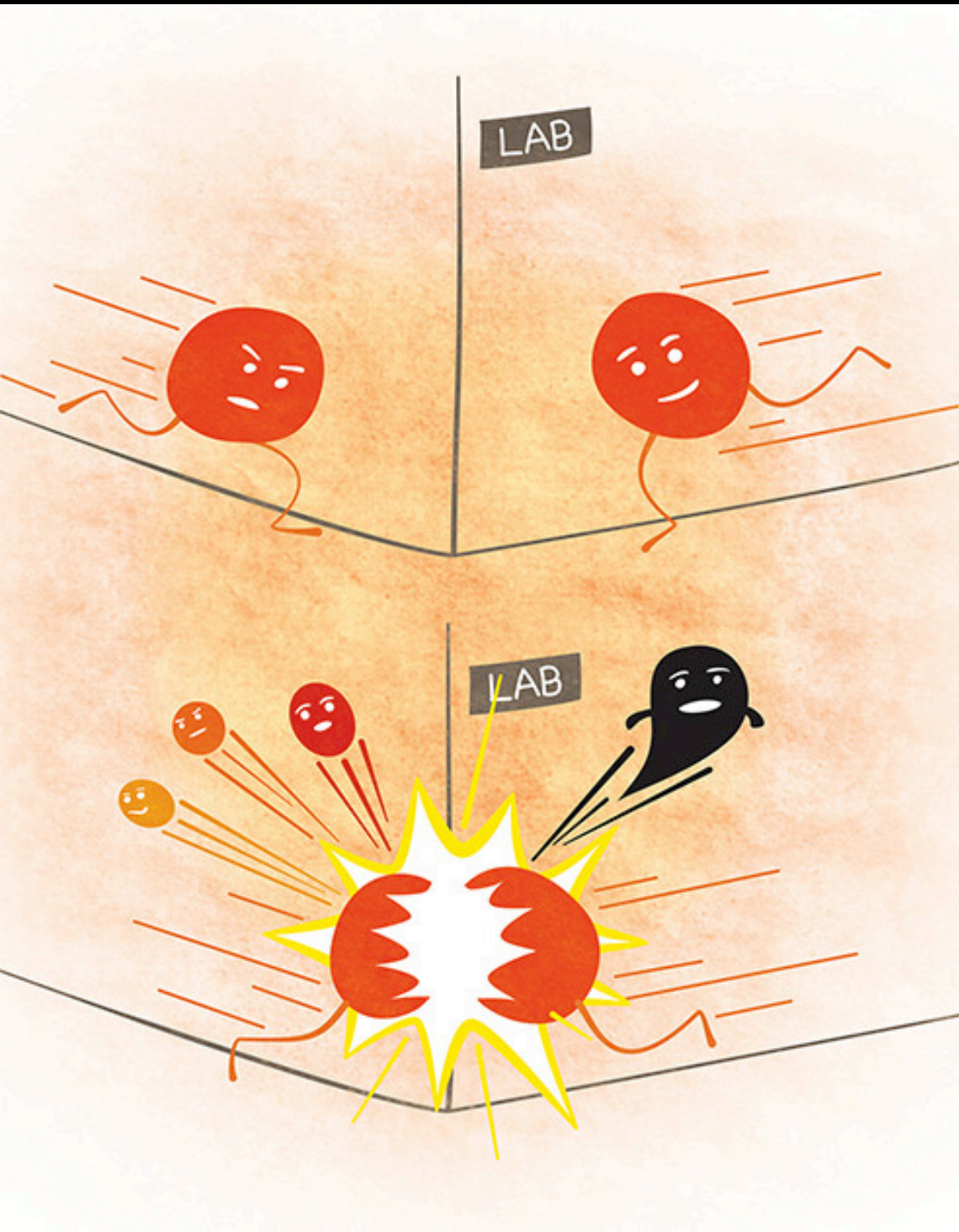


MOST OF THE MATTER IN THE UNIVERSE DOES NOT INTERACT (EMIT OR ABSORB) WITH LIGHT

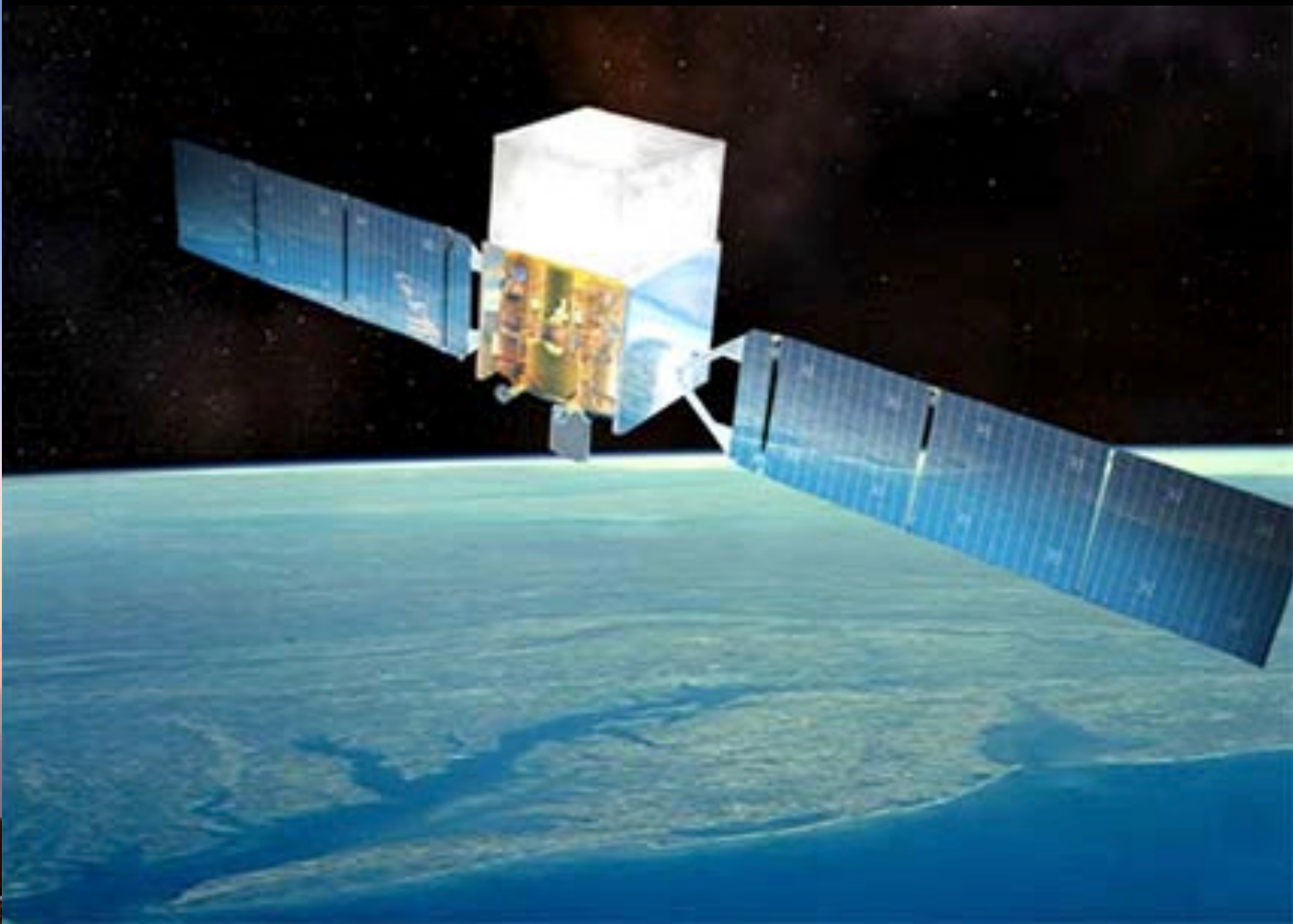


THREE WAYS TO LOOK FOR DARK MATTER

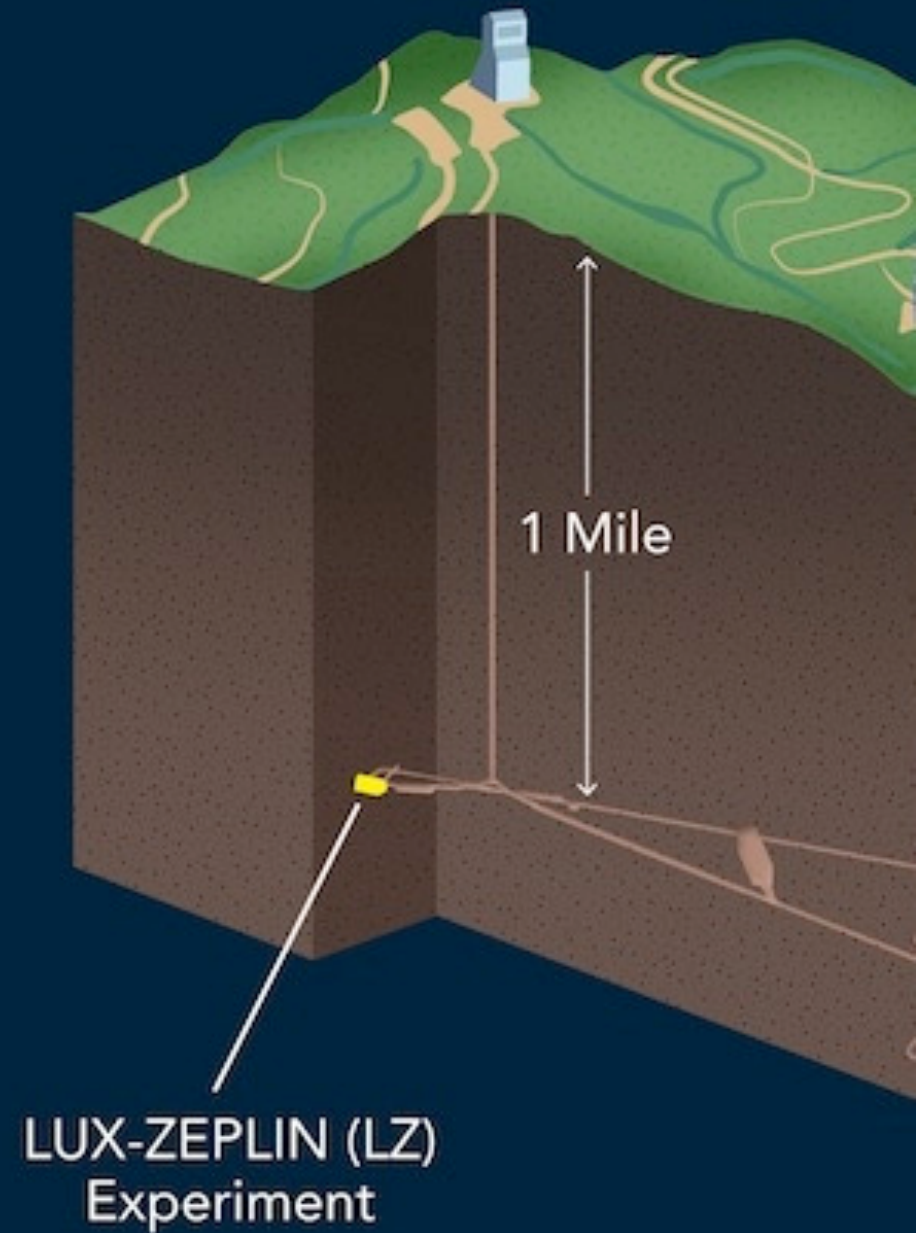
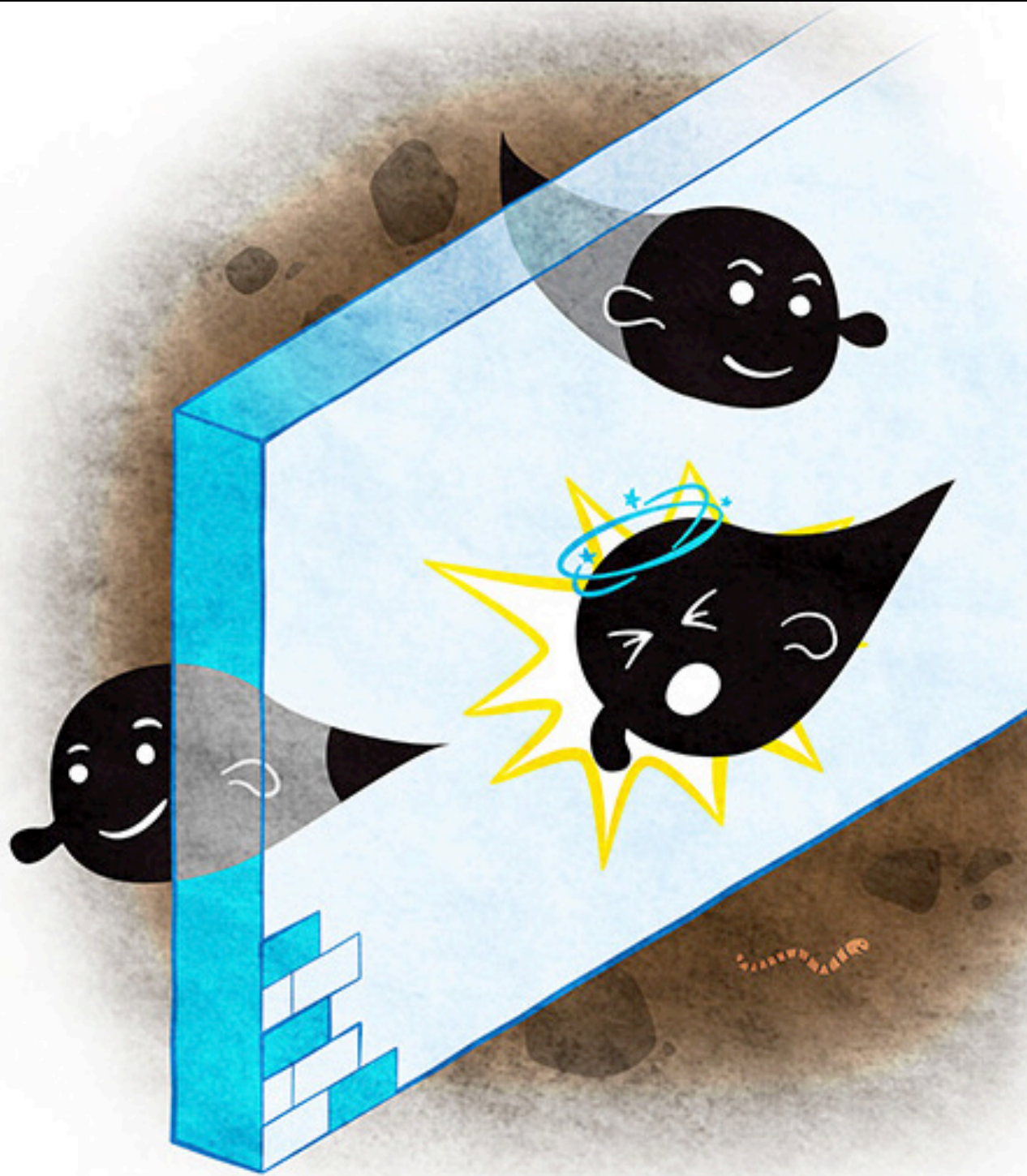
1. ENCOUNTERS IN THE LAB

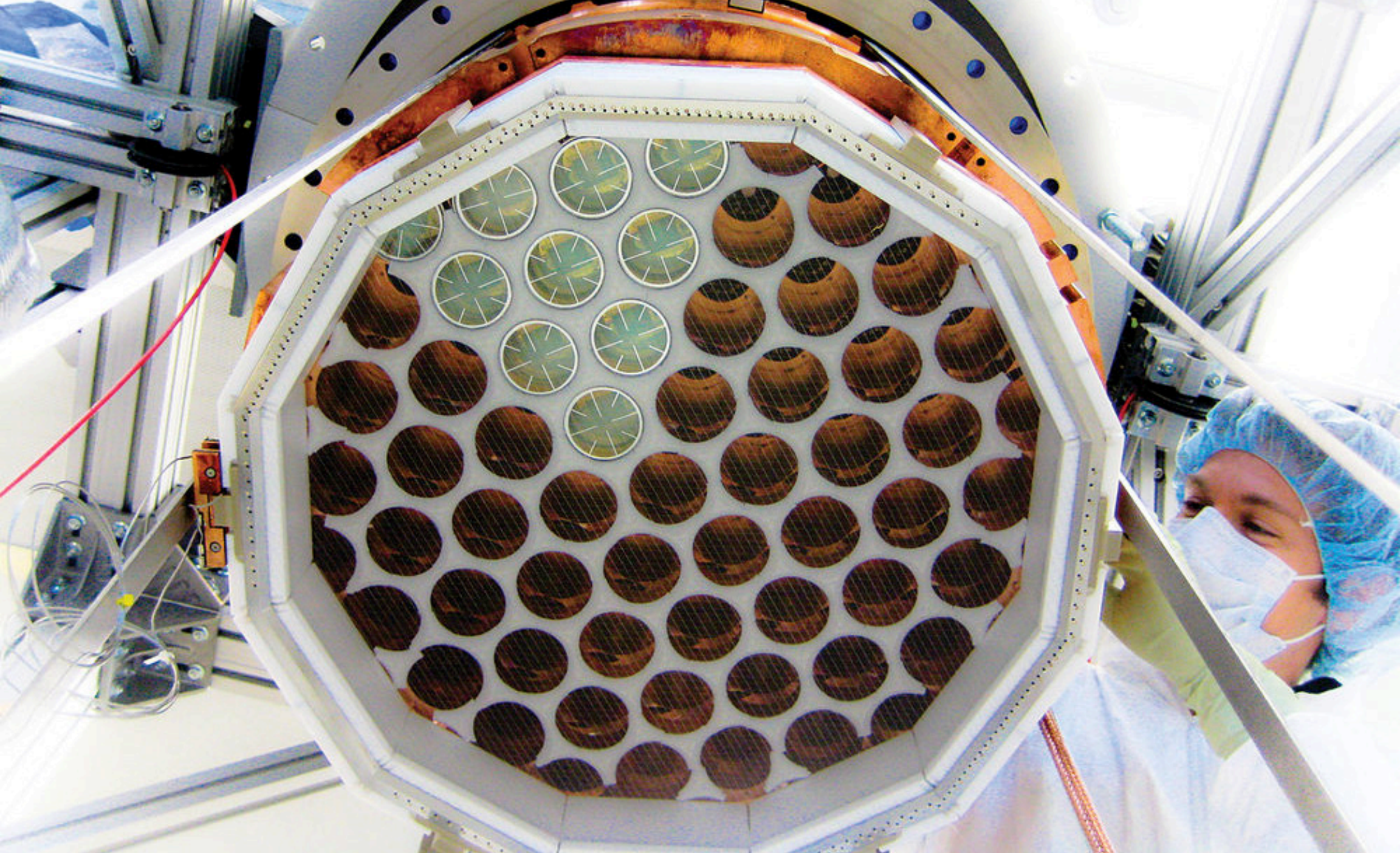


2. TRACING KNOCKOUTS IN SPACE



3. GOING BUMP UNDERGROUND!





HOW MUCH DATA ARE WE TALKING ABOUT?

THE FERMI GAMMA-RAY SPACE TELESCOPE

Launched on June 11, 2008

- from Kennedy Space Center
- launch vehicle: Delta II Heavy
- circular orbit, 565 km altitude, 25.6 degrees inclination

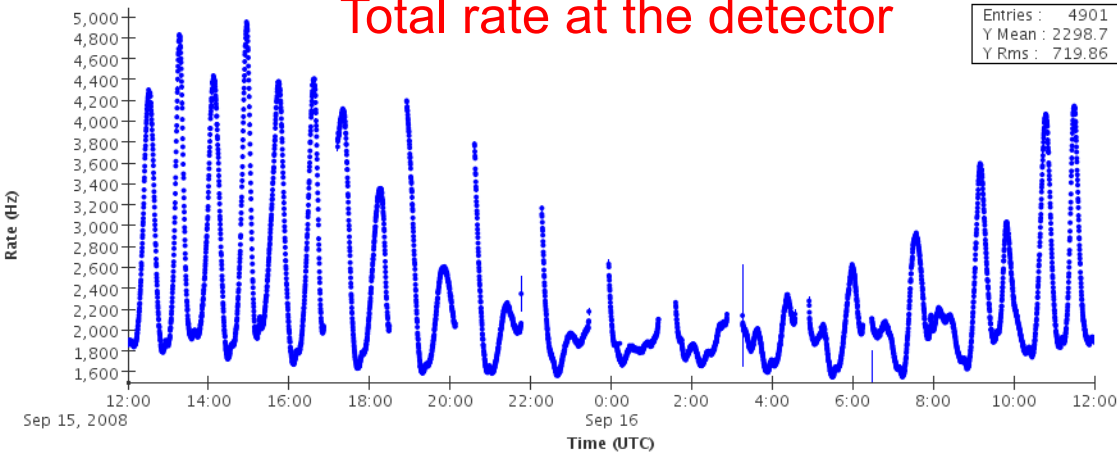
LAT milestones:

- 82,300 orbits to date
- 800 billion event triggers
- 179 billion evts downlinked
- 1.8 billion photons available in the public database

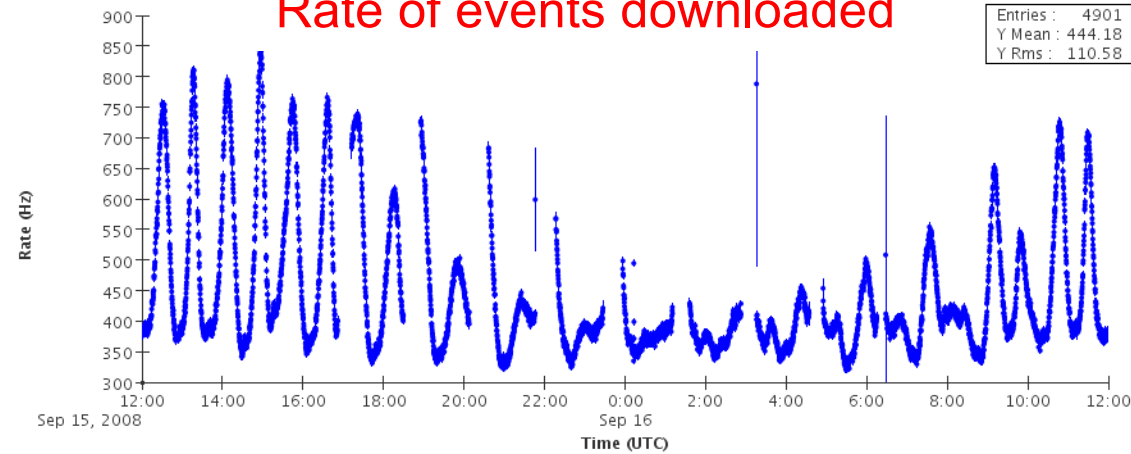


DATA RATES IN LOW EARTH ORBIT

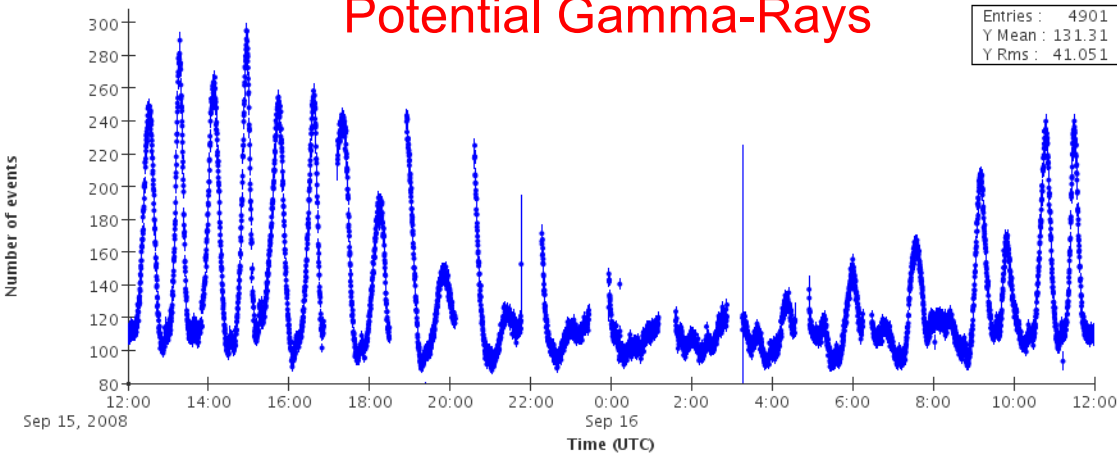
Total rate at the detector



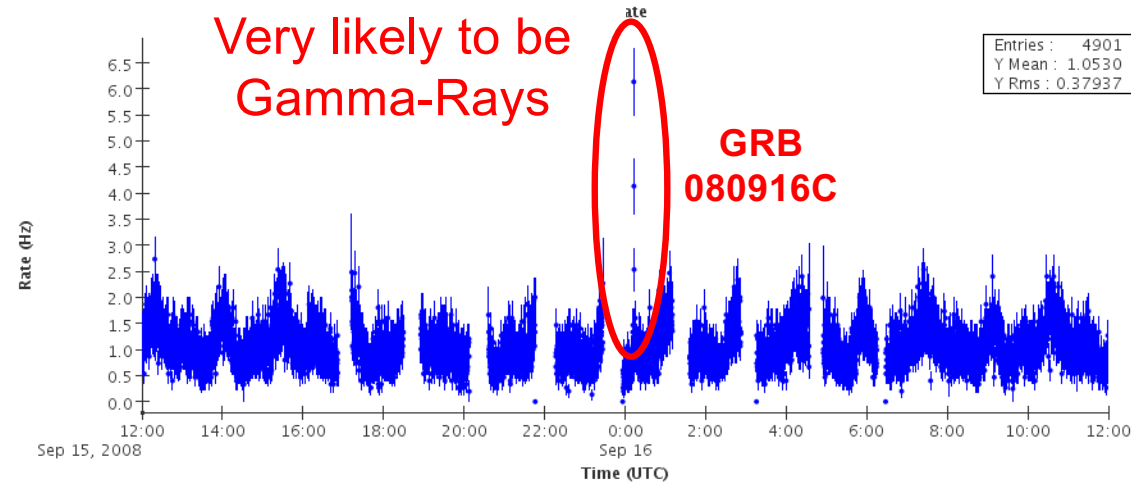
Rate of events downloaded



Potential Gamma-Rays



Very likely to be Gamma-Rays



- **0.3 PB OF DATA EACH YEAR, OR 5 PB TOTAL SO FAR**
- **LESS THAN 1% OF THE DATA IS A LIKELY GAMMA-RAY**

A SUCCESSFUL OPEN SCIENCE EXPERIENCE!

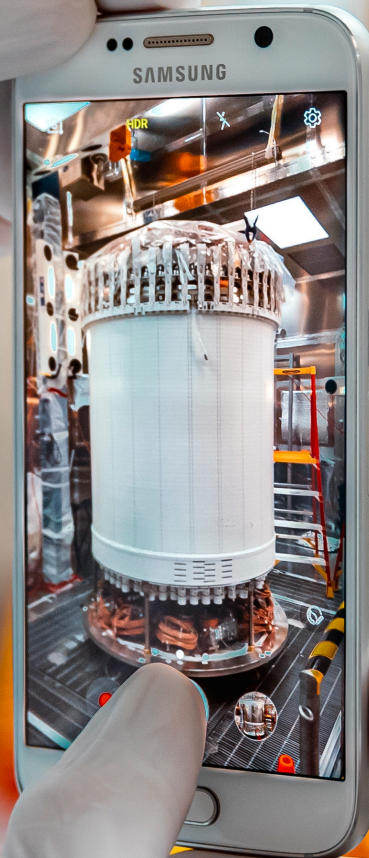
ALL FERMI DATA AND SOFTWARE ARE OPEN (SINCE 2009)



- **HOSTED @NASA/HEASARC**
- **INCLUDES: DATA, ANALYSIS SOFTWARE, INSTRUMENT RESPONSE, BACKGROUND...**
- **TREMENDOUS COMMUNITY ENGAGEMENT WITH OUR DATA**
- **90% OF SCIENTIFIC RETURN COMES FROM THE COMMUNITY**
- **REPRODUCIBILITY: QUALITY CONTROL OF ALL RESULTS**

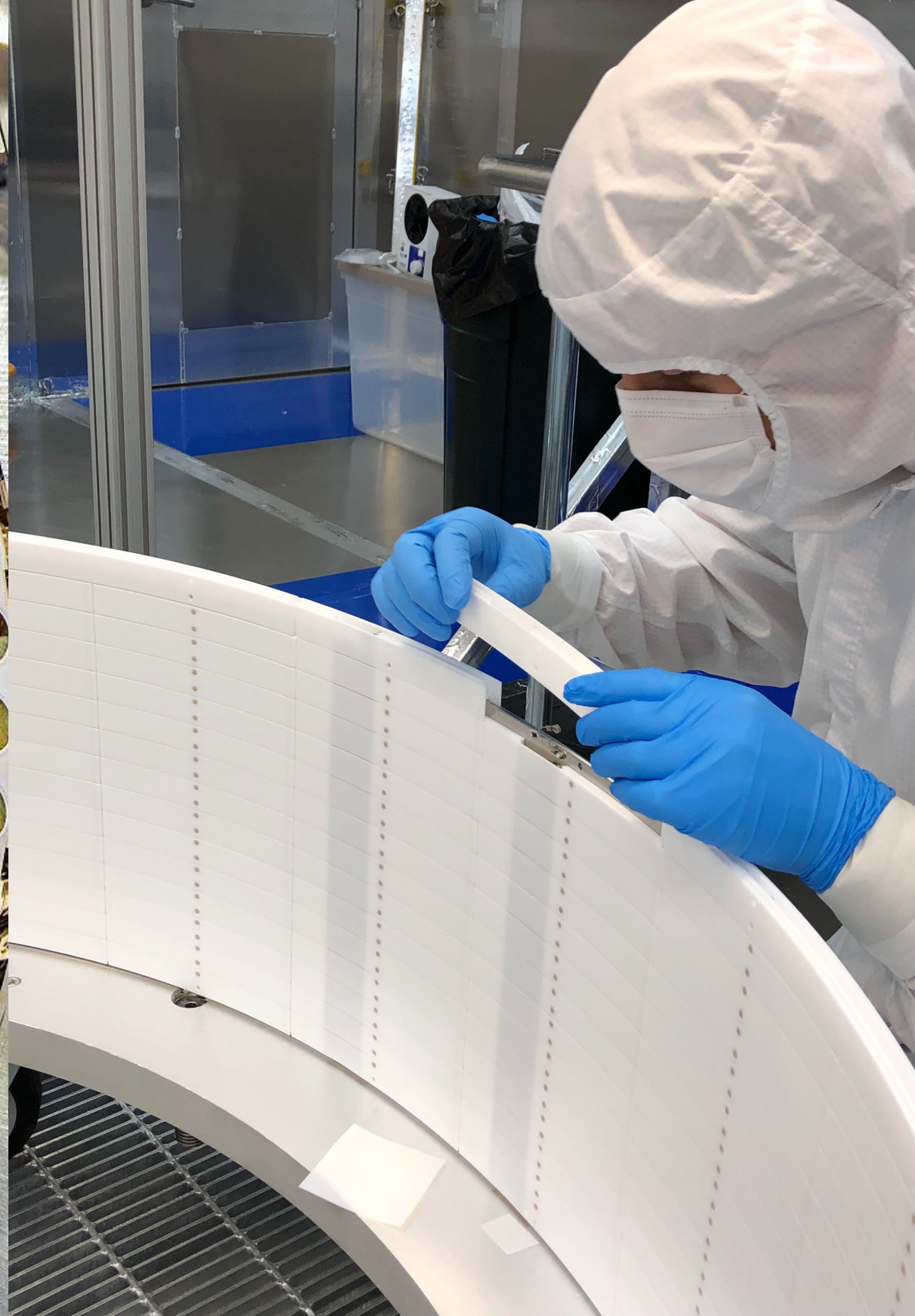
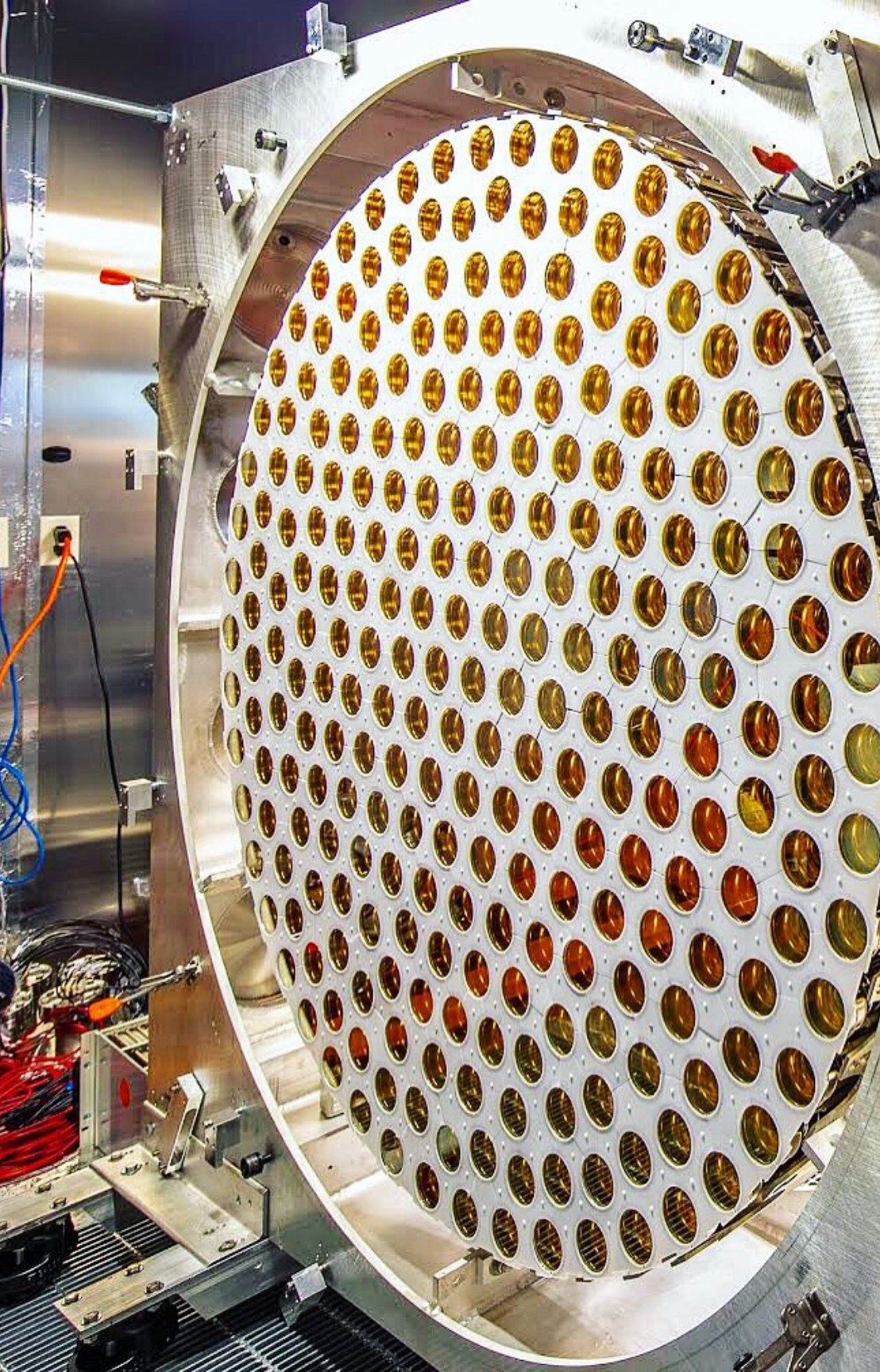
HOW TO BUILD A DIRECT DM DETECTION EXPERIMENT

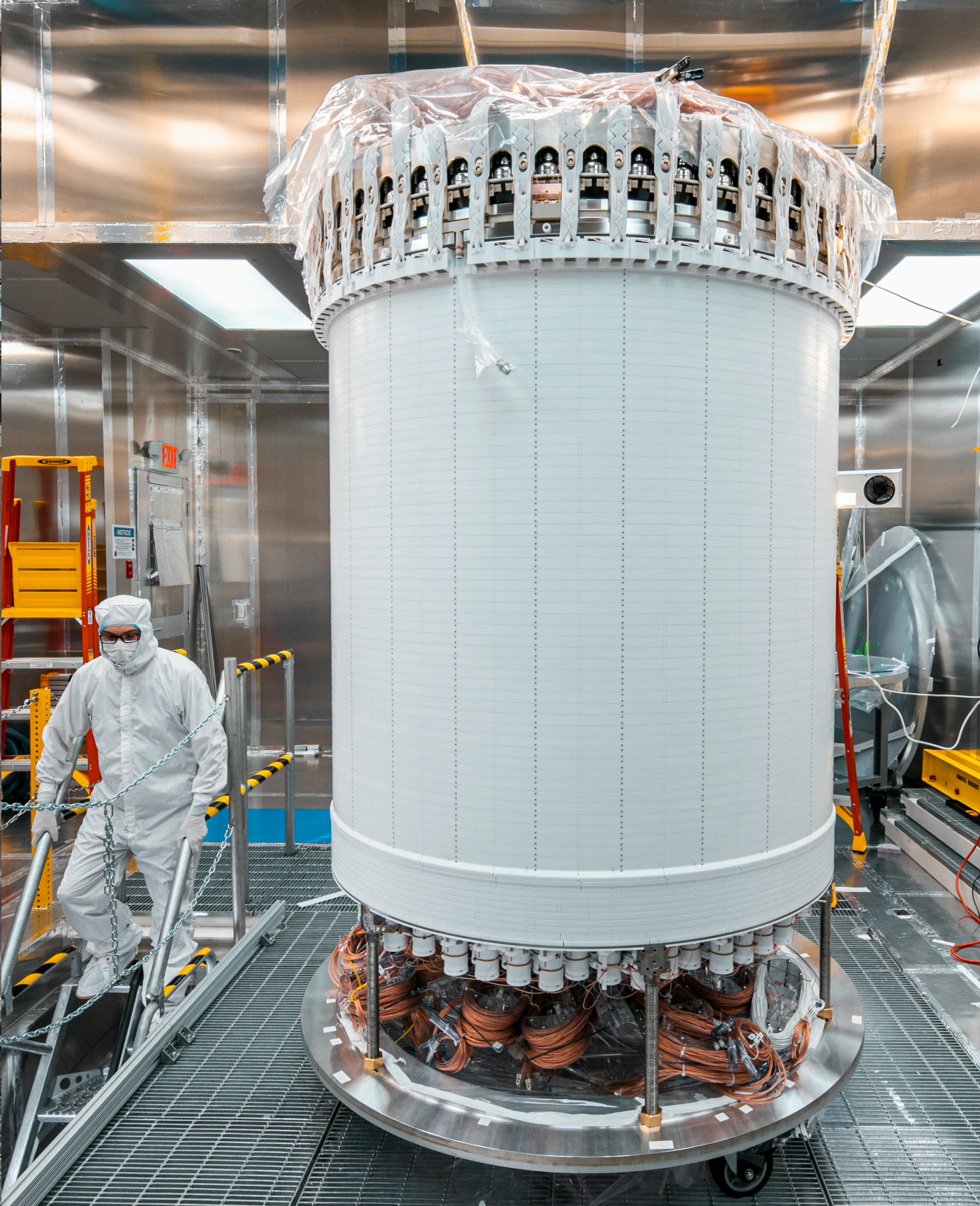
[=A BUCKET OF
LIQUID XENON]





PHOTOSENSORS







EVERYTHING IS RADIOACTIVE!!!



- **EVEN A MILE UNDERGROUND!**
- **WITH ULTRACLEAN MATERIALS [100,000 – 1MILLION TIMES CLEANER THAN "REGULAR"]**
- **DATA RATE \sim 50 COUNTS/SEC (THIS IS "ALL" BACKGROUND)**
- **OR... 1 BILLION COUNTS/YEAR**
- **OR... 1 PB DATA/YEAR**
- **5 PB TOTAL EXPECTED DATA 🤯**

DARK MATTER IS EXTREMELY ELUSIVE!

- **THERE ARE ~ 3 DM PARTICLES PER LITER ON EARTH**
- **STREAMING THROUGH THE EARTH AT 230 KM/SEC**
- **1 BILLION PARTICLES GO THROUGH LZ EVERY SECOND!**
- **OF THOSE DARK MATTER PARTICLES, WE ARE HOPING TO DETECT A HANDFUL PER YEAR (IF WE ARE LUCKY)!**
- **EXPECTED BACKGROUND INTERACTIONS: ~ 1 BILLION PER YEAR (OR 1 PB/YEAR OR 50/SECOND)!!!**

HOW DO WE FIND THOSE FEW PARTICLES?!?

NEEDLE IN A HAYSTACK



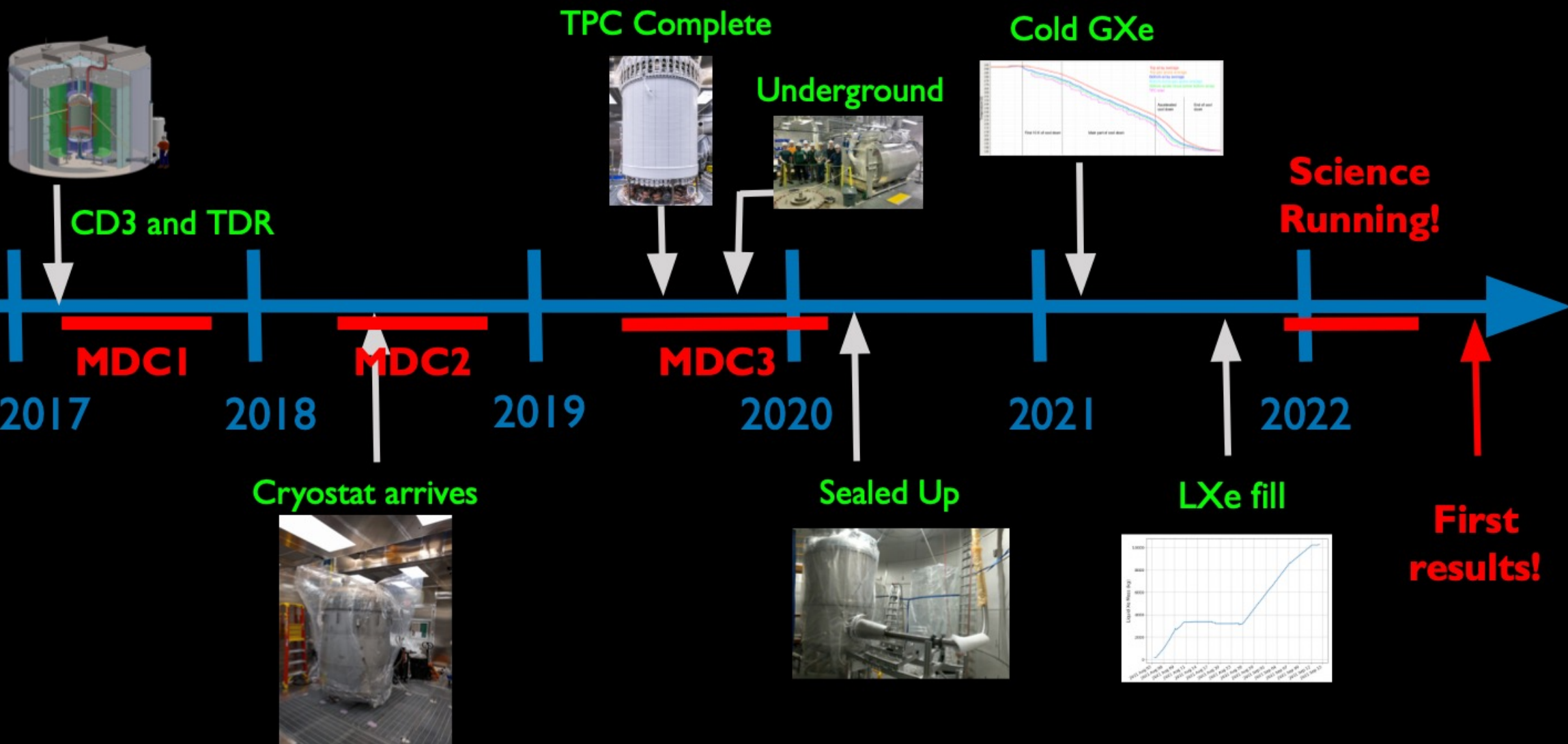
HOW DO WE FIND THOSE FEW PARTICLES?!?

NEEDLE IN A HAYSTACK: COMPUTING

- **AUTOMATIC 24/7 PROCESSING OF ALL DETECTOR DATA**
- **LARGE-SCALE SIMULATIONS WITH DETAILED MODELING**
- **EXTENSIVE CAMPAIGN(x3) OF MOCK DATA CHALLENGES**

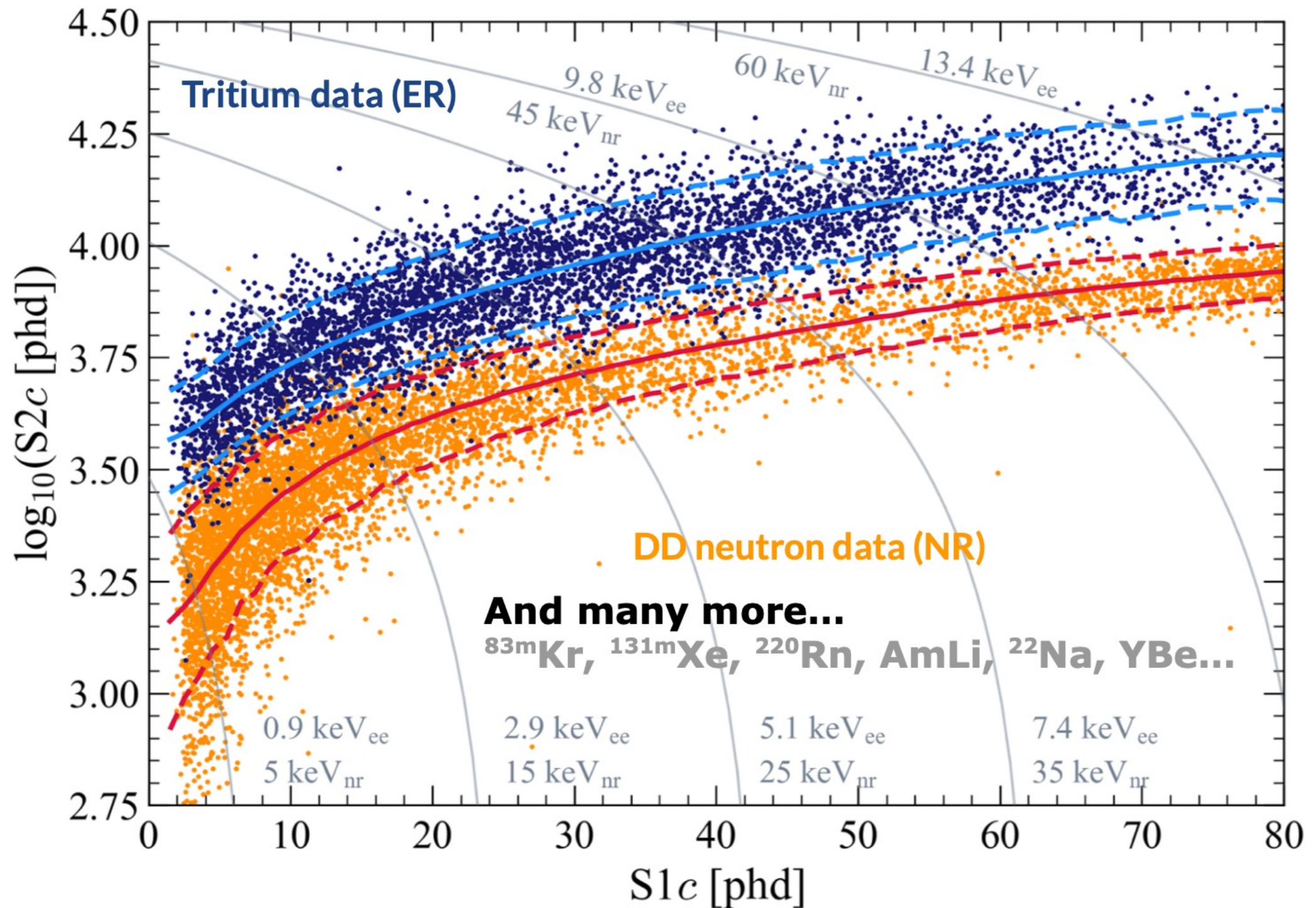


5 WEEKS FROM DATA TAKING TO RESULTS!

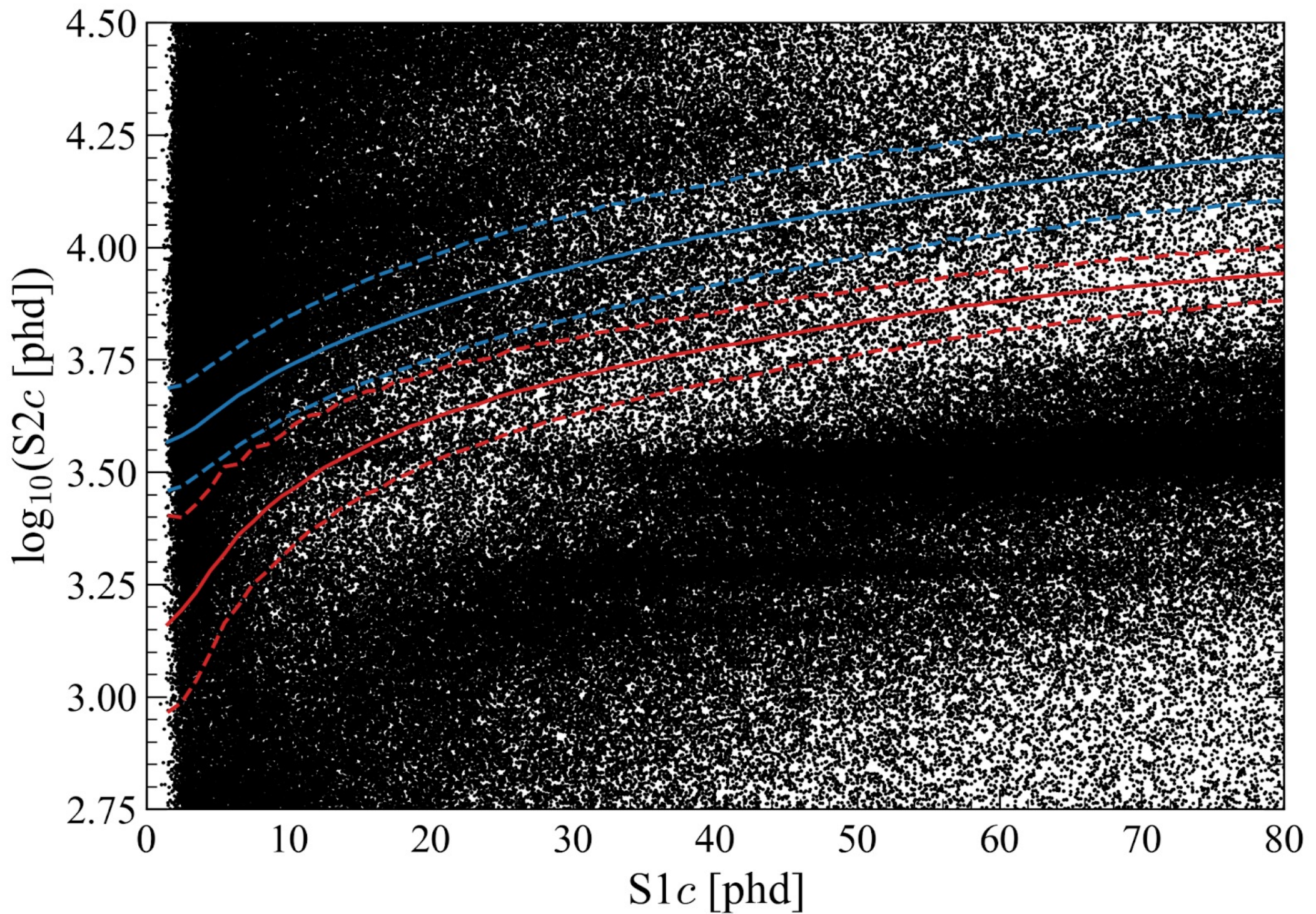


**ENABLED BY AN EXTENSIVE CAMPAIGN
OF MOCK DATA CHALLENGES (MDC)**

LZ DATA TAKING: CALIBRATIONS!



THE DATASET -> LZ: SCREAMS!!!



THE DATASET: EVENT SELECTIONS

All triggers

Time hold-offs

high rates of spurious instrumental activity, dominated by post-S2 hold-off (70% live fraction)

Low energy single scatters

$3 < S1c < 80$ phd, $S2c > 600$ phd ($10e^-$)

Pulse quality cuts

target accidental coincidence events

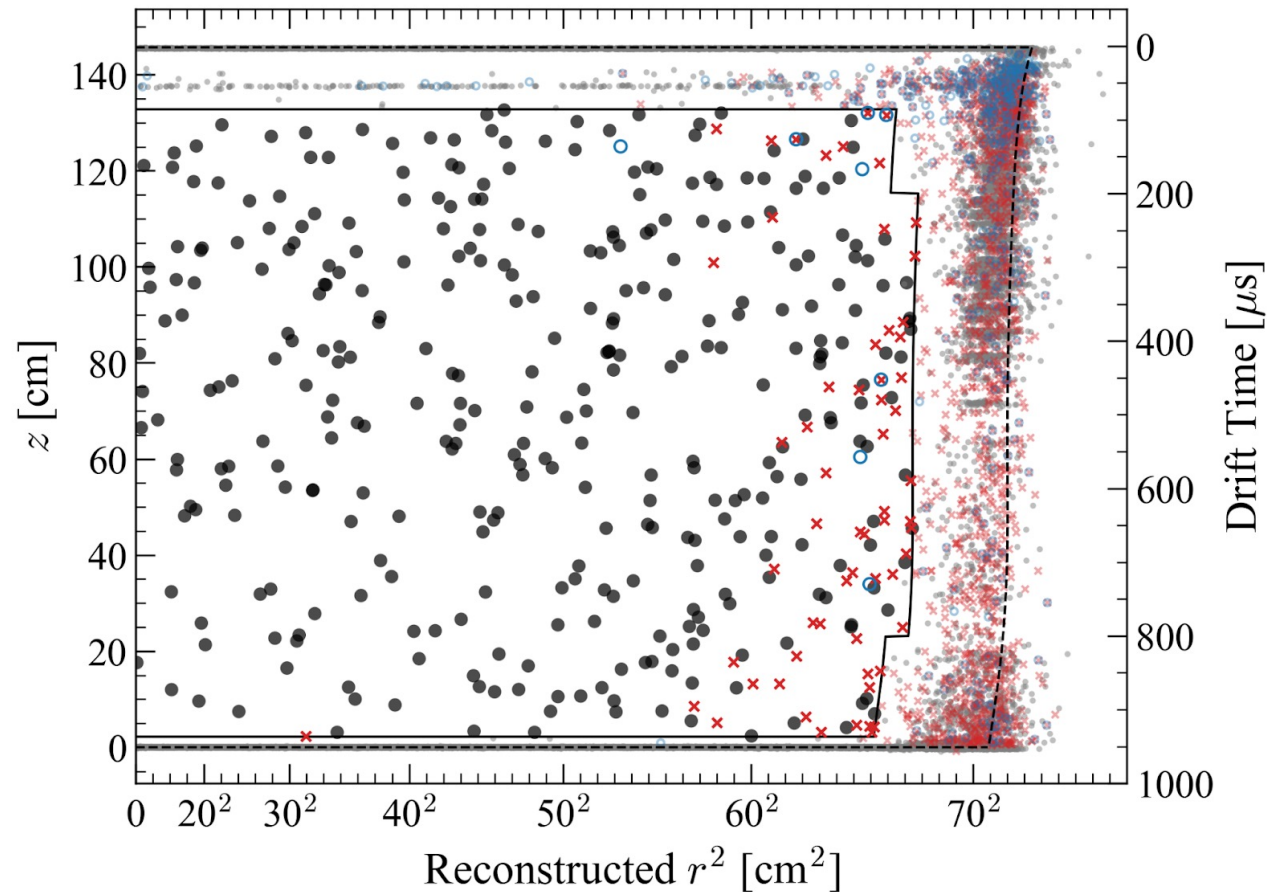
Fiducial volume

central 5.5 tonnes of LXe

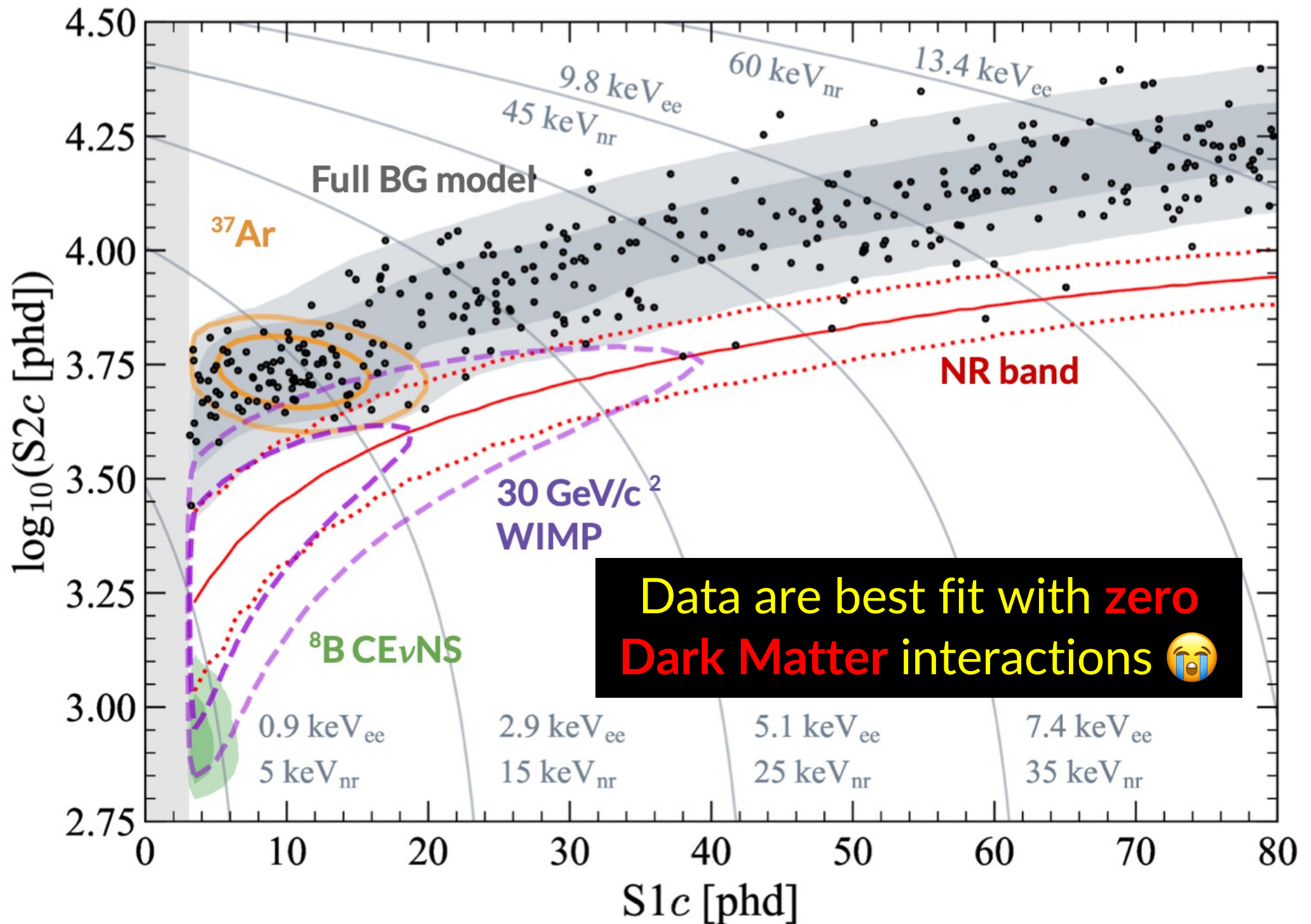
OD + Skin vetoes



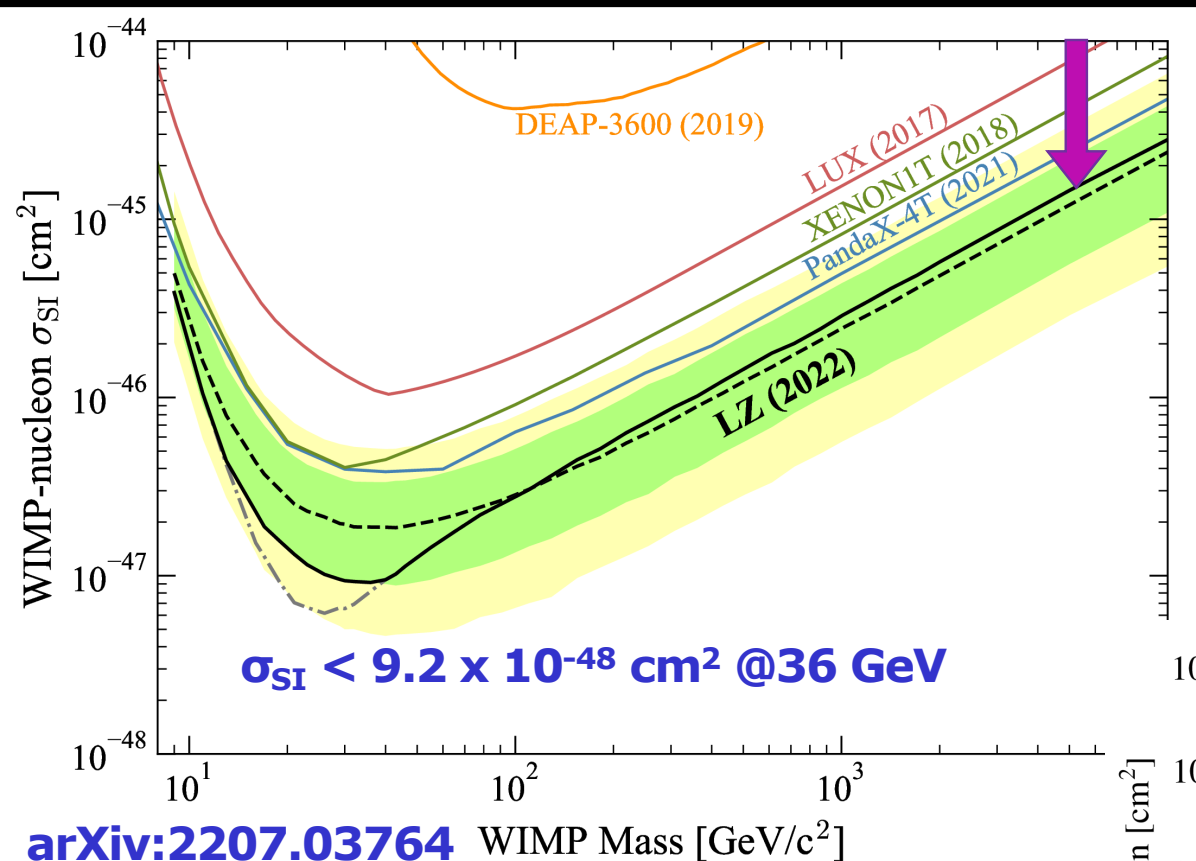
5.5 tonne inner volume



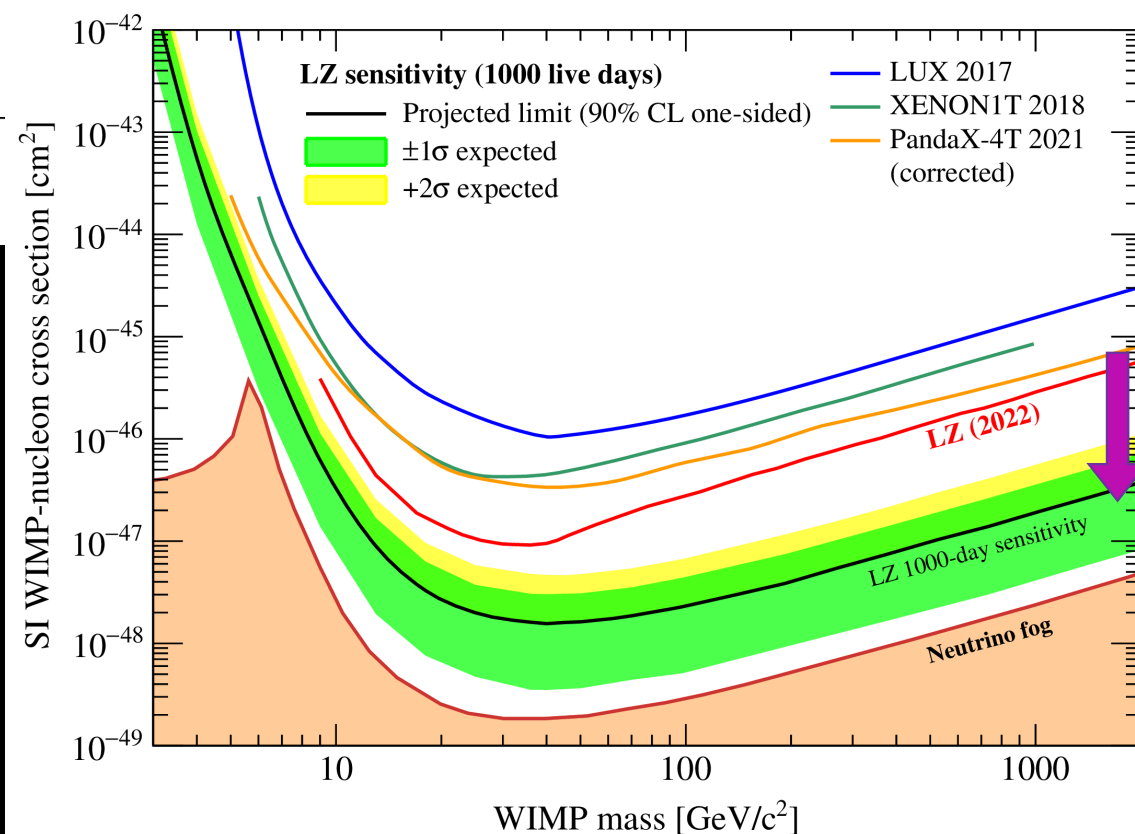
HEY DATASET, YOU CLEAN UP REAL NICE 🤪



FINDING NOTHING, BETTER THAN EVERYONE ELSE



NOT FINDING THE DARK
MATTER PARTICLE TELLS US
A LOT ABOUT THE DARK
MATTER PARTICLE

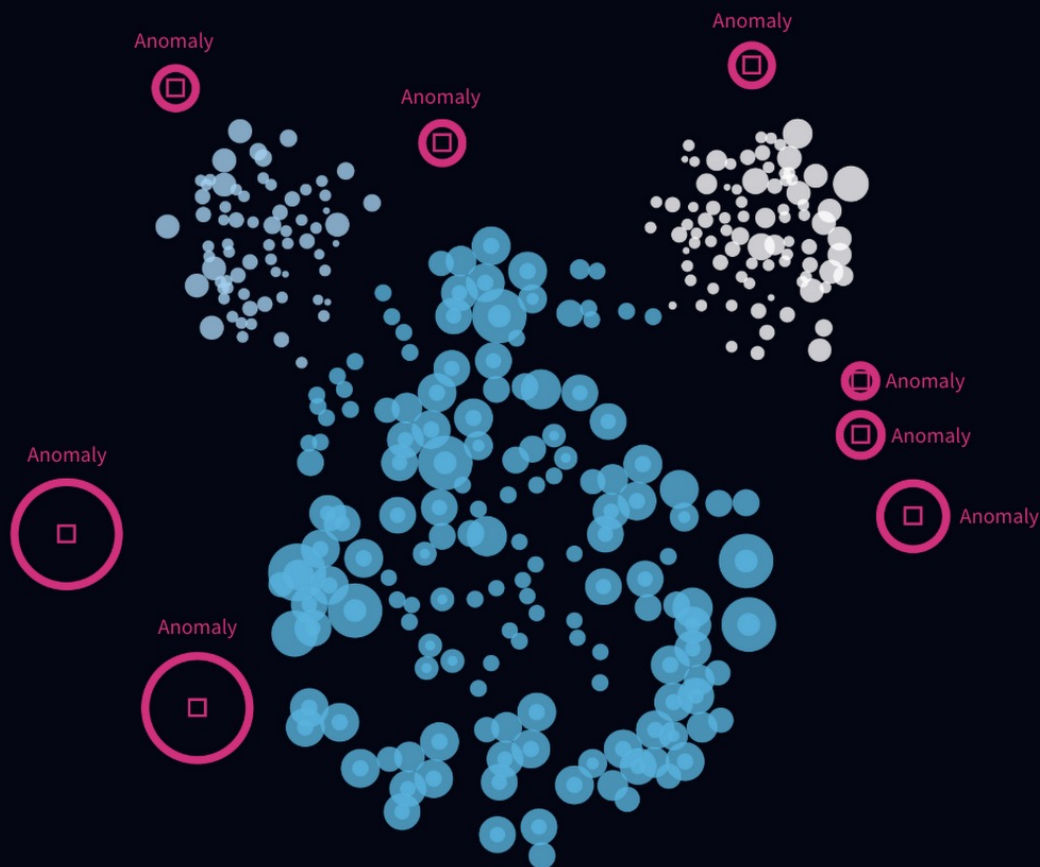


OK, WHAT'S NEXT?

LET'S WAIT SOME MORE!
AKA: LOTS. MORE. DATA

CAN WE USE SOME FANCY ALGORITHMS?

- **ISSUE: REJECT RARE/UNPHYSICAL DETECTOR EFFECTS**
- **APPROACH: ANOMALY DETECTION (CLUSTERING, VAE)**
- **SEE ALSO: RYAN HUMBLE & TYLER ANDERSON'S TALKS**



Anomaly detection using clustering
(diagram from Google Cloud blog)

Clustering of LZ data with UMAP +
DBSCAN (M. Arthurs' dissertation)



EXCEEDINGLY (?) REDUNDANT ARCHITECTURE

Fully redundant design (2 complete “live” copies of all data + either data center can perform full simulations/data processing)!

Upside: exceptional level of risk mitigation

Downside: diverging architectures, plus a policy nightmare

We've gotten VERY good at SW portability

Phenomenal training opportunity for early career scientists (->retention challenge 🤖)



US Data Center:

- Prompt Processing
- Long-term Archive
- Supercomputers!



UK Data Center:

- Data Reprocessing
- Sims Production
- Distributed CPUs!



SUPPORTING REALTIME SCIENCE @ NERSC

PROCESS DETECTOR DATA EARLY AND OFTEN:

- Automatic prompt-processing (reconstruction) upon data reception
- Redundant capabilities to reprocess/simulate multiple times based on calibration/analysis results (rerun 1 year of data in 1 month)

TIME IS OF THE ESSENCE! RAPID TURNAROUND:

- Very limited computing resources are available at SURF (RAID array for storage and “first look” quality monitoring - See S. Luitz' Talk)
- Full-scale data quality monitoring/detector health assessment happen at NERSC. Quasi-real-time analysis feedback during commissioning

EXTENSIVE USE OF SPIN-BASED SERVICES:

- Supporting both production/monitoring tools and user/data access!
- SPIN: "Containers-as-a-Service (CaaS)" platform @NERSC

CONTAINERS-AS-A-SERVICE PLATFORM

SPIN IS A CONTAINER-BASED PLATFORM AT NERSC SUPPORTING THE DEPLOYMENT OF SCIENCE GATEWAYS, WORKFLOW MANAGERS, AND OTHER WEB SERVICES. IT SEAMLESSLY INTEGRATES NERSC COMPUTE/STORAGE.

Build

images on your laptop with your custom software, and when they run reliably, ...



Ship

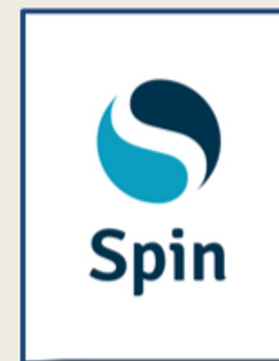
them to a registry for version control and safekeeping

- DockerHub: share with the public
- NERSC: keep private to your project

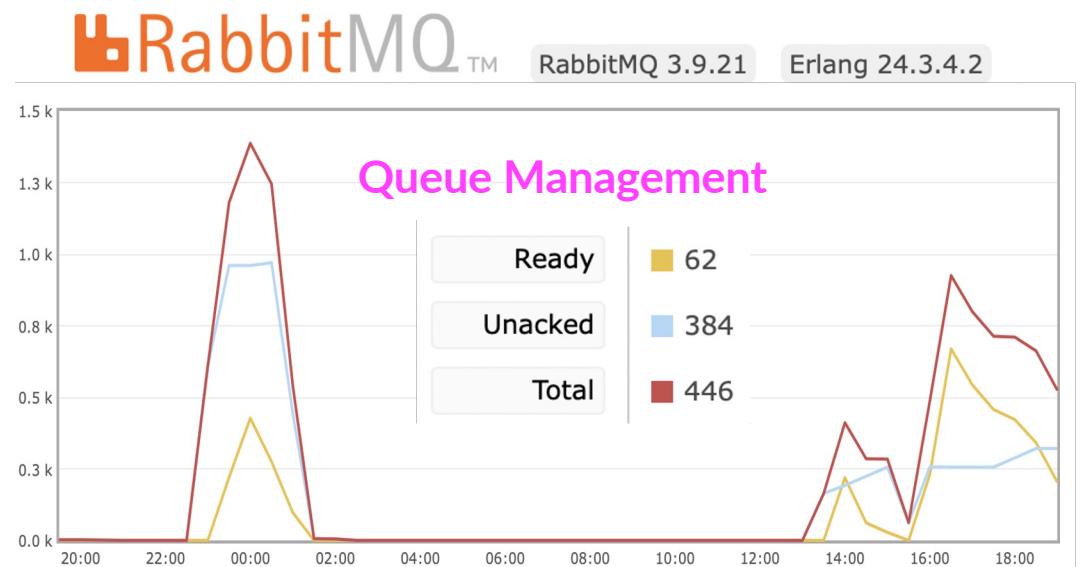
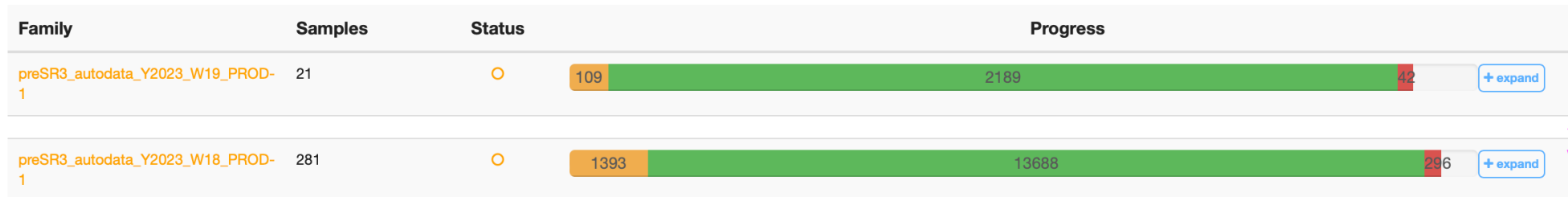
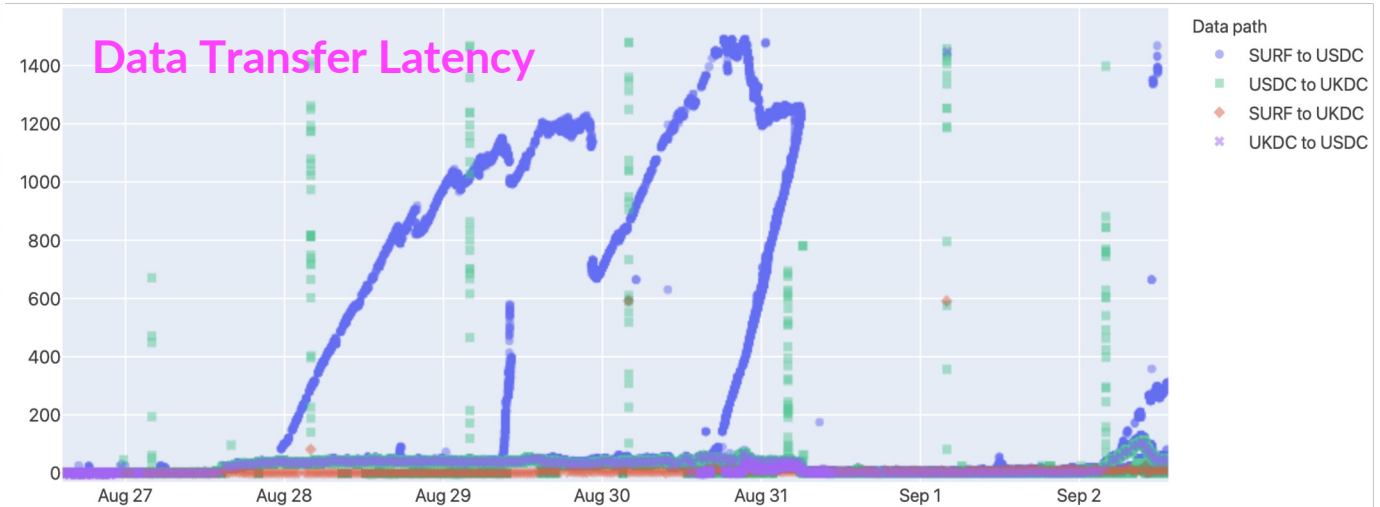
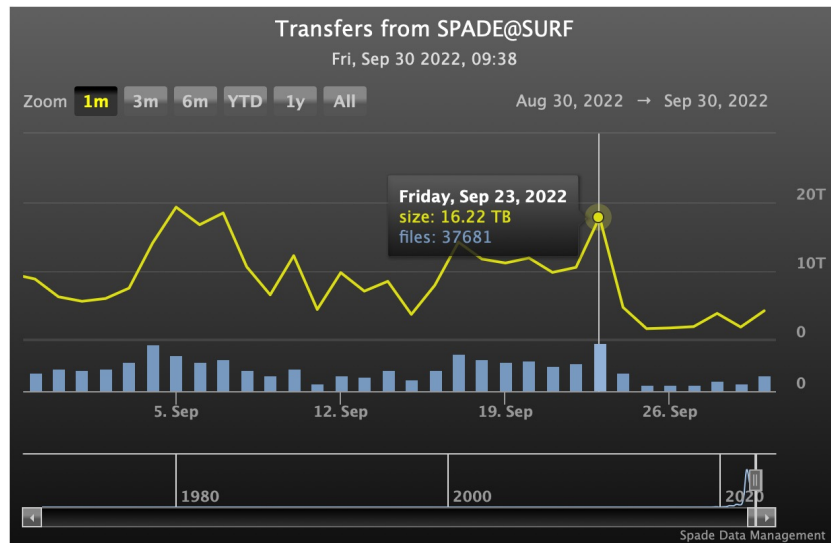


Run

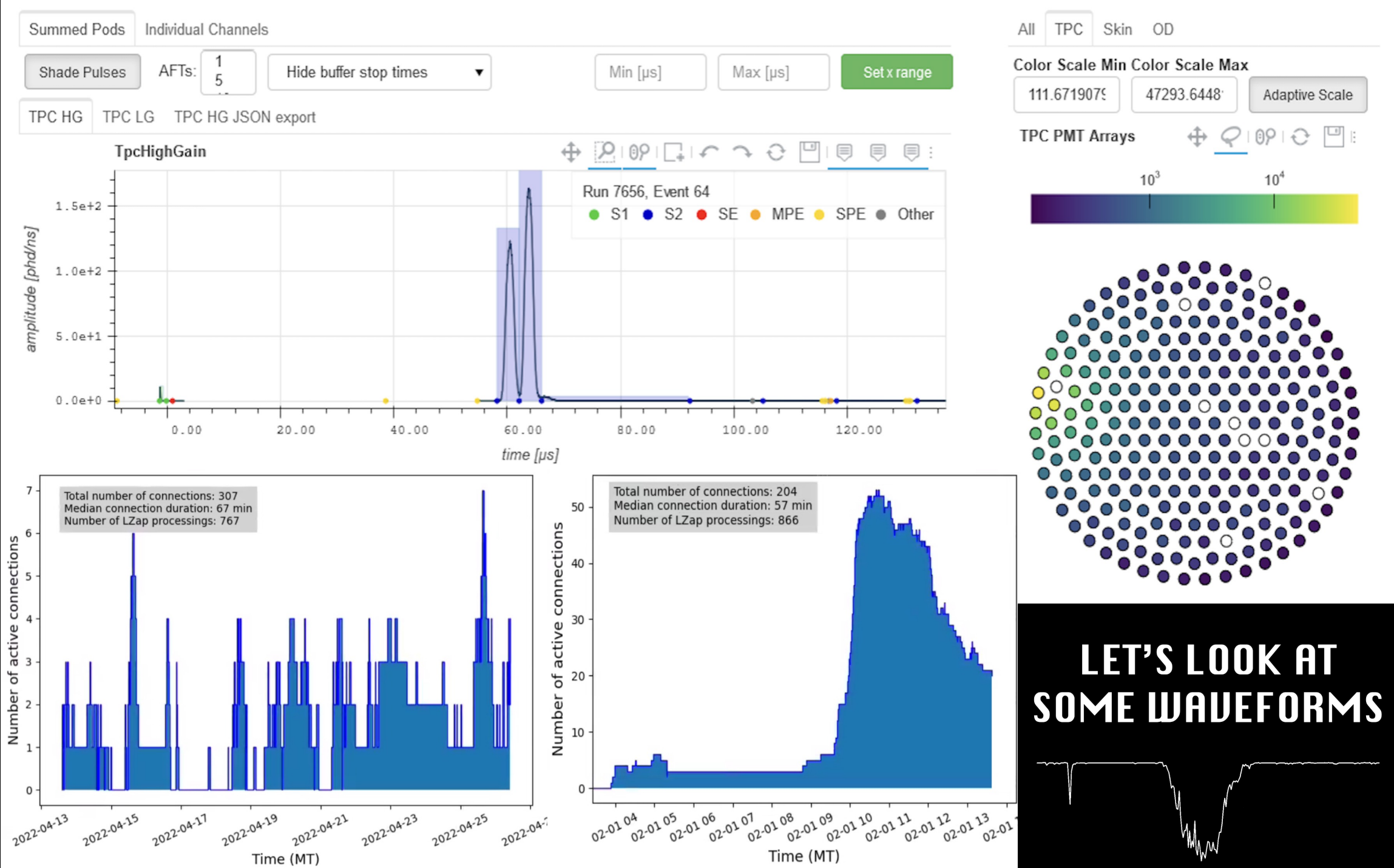
your workloads



MONITORING DATA MOVEMENT & PROCESSING



OFFLINE EVENT VIEWER (VERY POPULAR IN LZ!)



LZ's EXPERIENCE WITH SPIN

LZ's Offline Web Services

This is the main page served by lux-zeplin.lbl.gov

Services hosted by LZ at this domain:

- [SPADE Monitor](#)
- [Prompt Processing Monitor](#)
- [Offline Run Tables](#)
- [LZ Event Viewer](#)
- [PREM: Offline Data Monitor](#)

Services hosted elsewhere (login required):

- [Online page hosted at SURF](#)
- [Twiki Document Repository](#)
- [Google Drive Folder](#)
- [LZ Code Repository in GitLab](#)
- [Offline Software Documentation](#)

Encrypted LZ NGINX Deployment ([logout](#))

PROS:

- Ease of use & excellent (hands-on!) training
- Integration with NERSC compute and storage

CHALLENGES:

- Reliability/Resiliency across multiple NERSC resources
- Scalability of underlying infrastructure (isolate DEV instance, DB optimization)
- Policy evolution (AKA: "The Devil is in the Details")

RELIABILITY, RESILIENCY, ROBUSTNESS

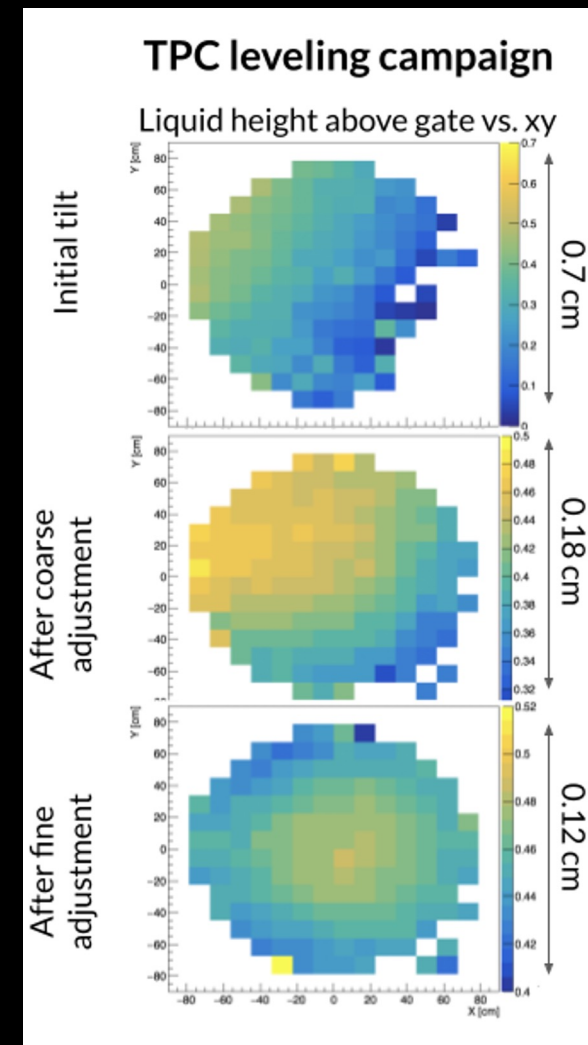
DOWNTIME IS EXPENSIVE!

- Defensive Engineering + Reputation with Science partners
- We use so many parts of NERSC, that downtime or degradation anywhere (DTNs, CFS, Slurm, SPIN, etc.) impacts entire workflow

IMPACT ON COMMISSIONING, OPERATIONS, CALIBRATIONS, DETECTOR HEALTH

- Example: TPC leveling campaign of 2021
- SURF "underground week": Mon-Thu
- Scheduled Cori outage on Wed of the same week
- Analyze data every night during TPC leveling
- NERSC kept us running on backup system

NON-DEGRADED SYSTEM AVAILABILITY IS TOP CONCERN FOR REALTIME COMPUTING



A TEAM EFFORT! THE LZ COLLABORATION (PRE-PANDEMIC PHOTO AT SURF)



Thanks to our
sponsors and
participating
institutions!



U.S. Department of Energy
Office of Science



Science and
Technology
Facilities Council

FCT

Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA



ibS Institute for
Basic Science