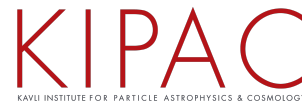




The Real-Time Data Workflow of LZ Dark Matter Experiment at NERSC

Maris Arthurs on behalf of LZ collaboration
CHEP 2024, Krakow

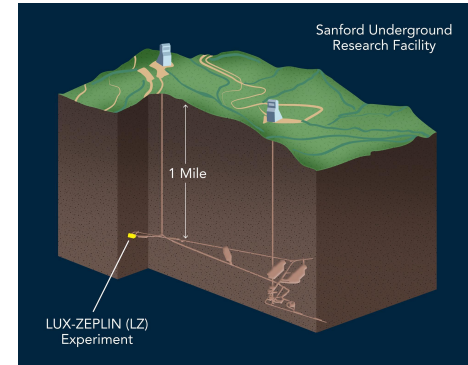
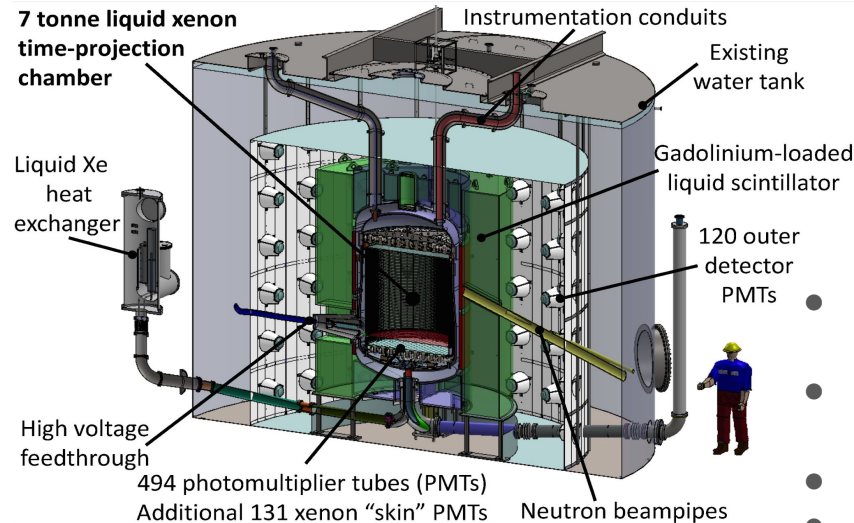
Oct 20, 2024



NATIONAL
ACCELERATOR
LABORATORY

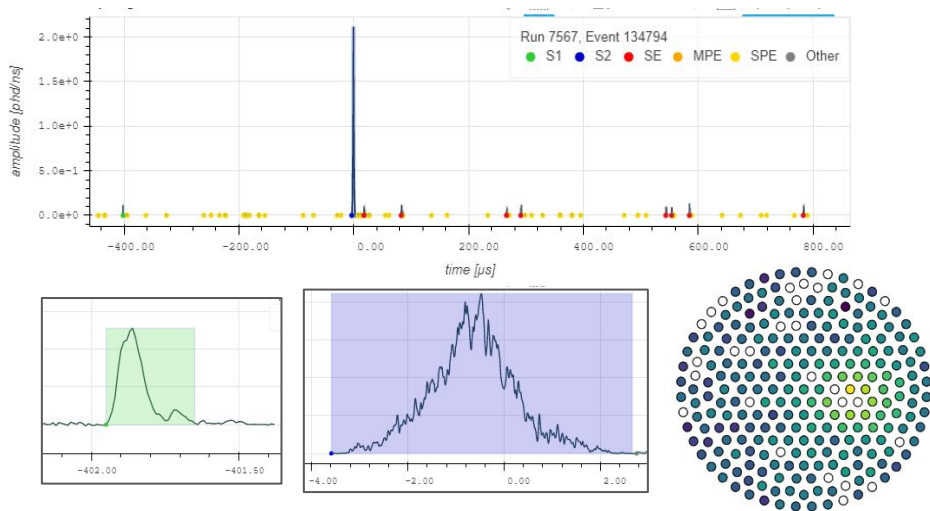
The LUX-ZEPLIN (LZ) Experiment

- LZ is designed for **direct detection** of **WIMP dark matter** interactions with xenon atoms
- Located about a mile underground at SURF, South Dakota



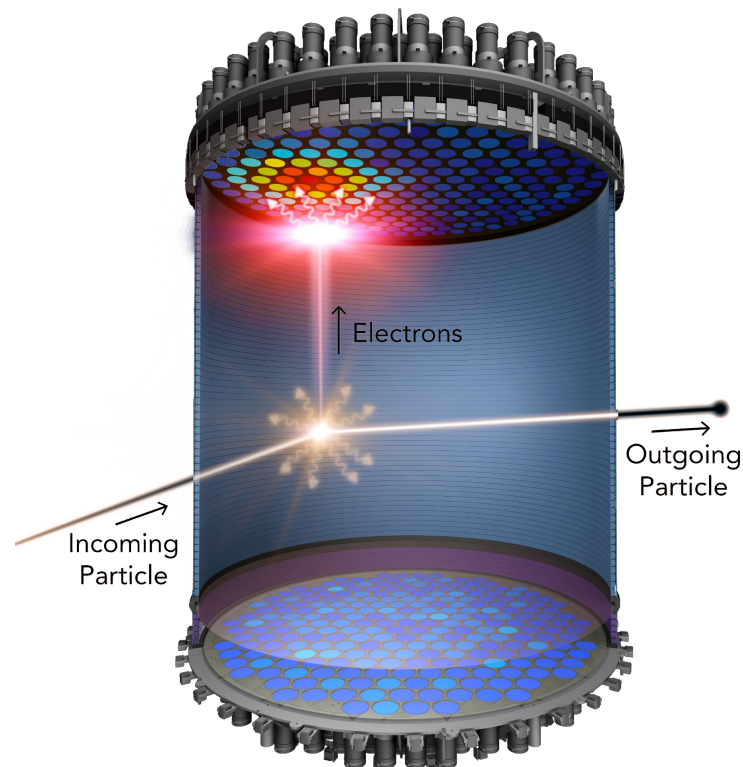
- The heart of LZ detector is a **7-tonne liquid xenon TPC**
- Detector commissioning finished late 2021
- 1st science results - July 2022
- WS 2024 - **World-leading WIMP sensitivity**

The LZ TPC

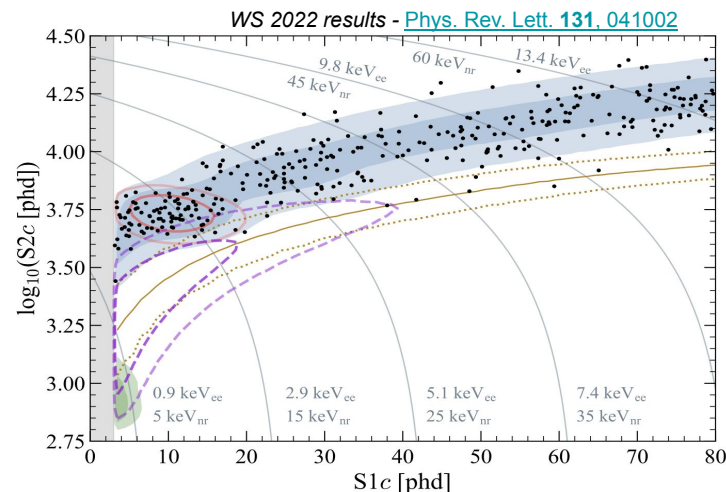
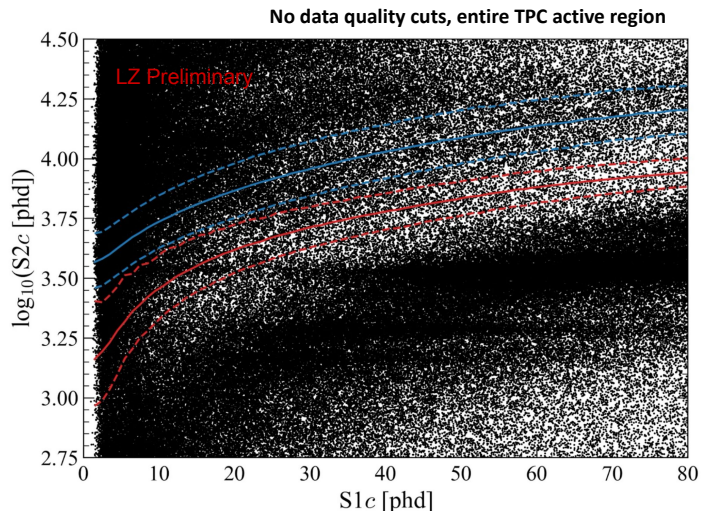


Dual Phase TPC Detector

- Primary scintillation (light) \rightarrow S1
- Secondary scintillation (from charge) \rightarrow S2
- Radial position from top PMT array S2 pattern
- Z position determined from the drift time
- S2/S1 ratio enables background discrimination



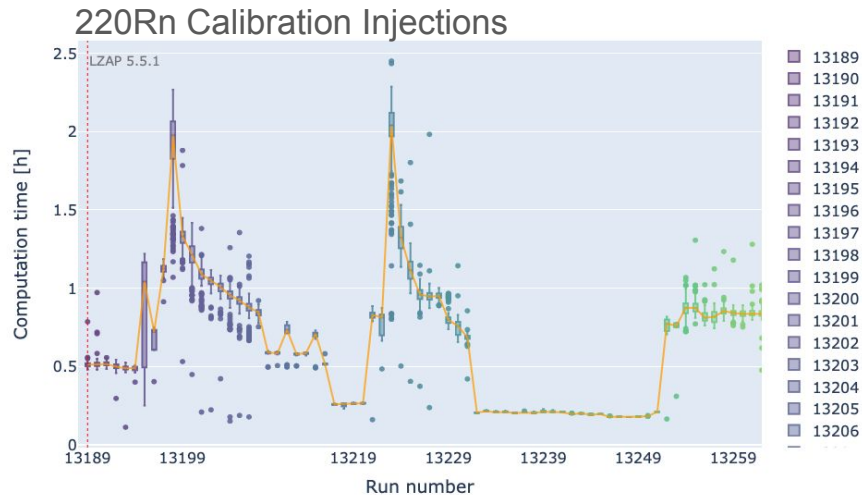
Rare-Event Search – Needle in a Haystack!



- About a billion WIMPs go through LZ per second
 - Hoping to detect a **few WIMPs per year**
 - We need to **collect a lot of data!** But that comes with a price
 - *Need to keep the background rates as low as possible!*
- Expect about **1B background interaction per year**
 - About 50 counts per second
 - 1PB Data per year

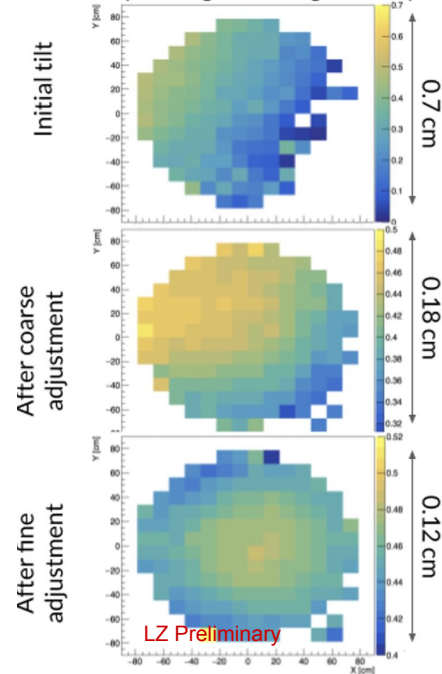
Need for Prompt and Real Time Computing

- Need **minimal downtime** and **fast turnaround time**
 - Impact on commissioning of the detector, operations, calibration, detector health, and data quality monitoring
 - Leveling of the TPC during commissioning
 - Prompt processing of calibration data for analysis



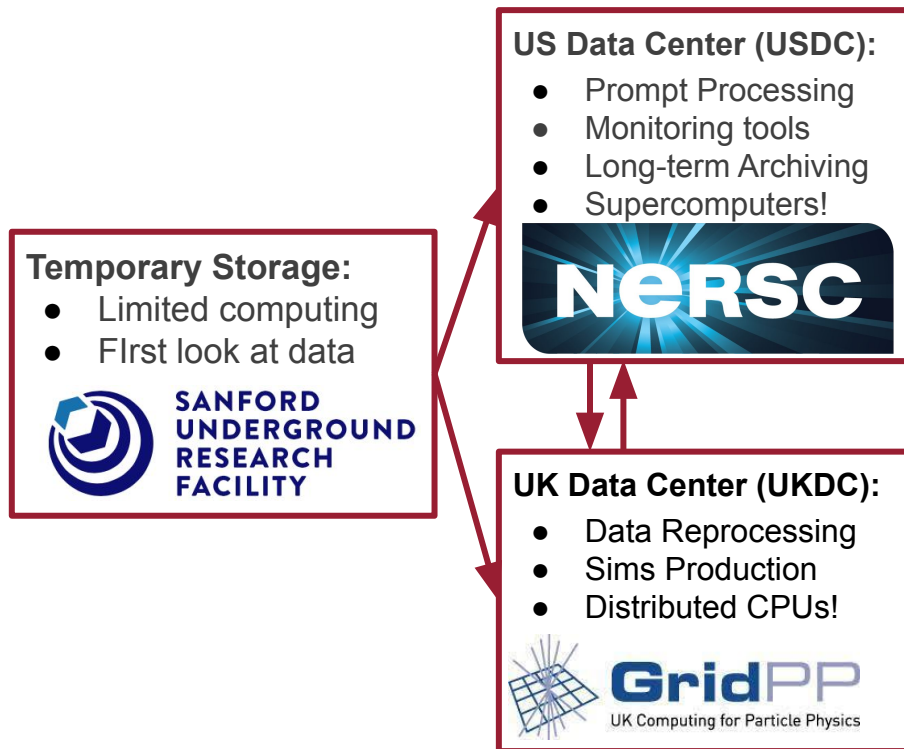
TPC leveling campaign

Liquid height above gate vs. xy



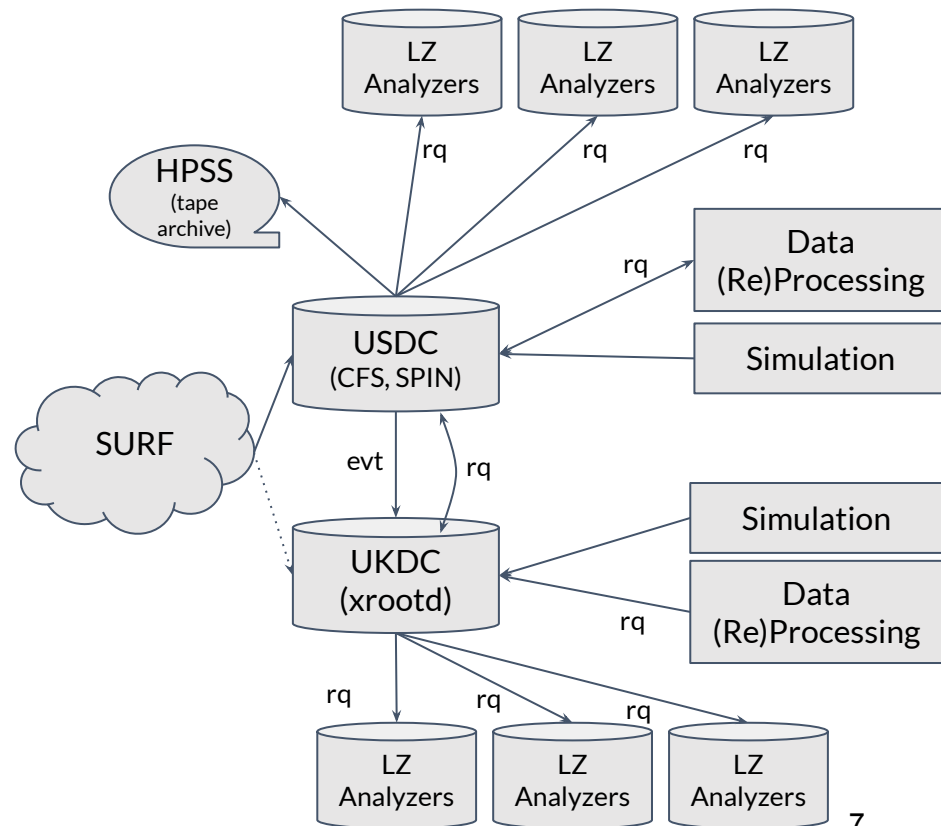
LZ Offline Computing

- Need robust computing infrastructure
 - Rapid turnaround time – **Automatic movement and processing (@NERSC)** of detector data
 - Large scale simulations
 - Data quality and Monitoring tools
- Fully Redundant design
 - **2 live copies of data** at US and UK data centers, 1 archived at US
 - Data rate is 3PB per year including raw, reconstructed and calibration data



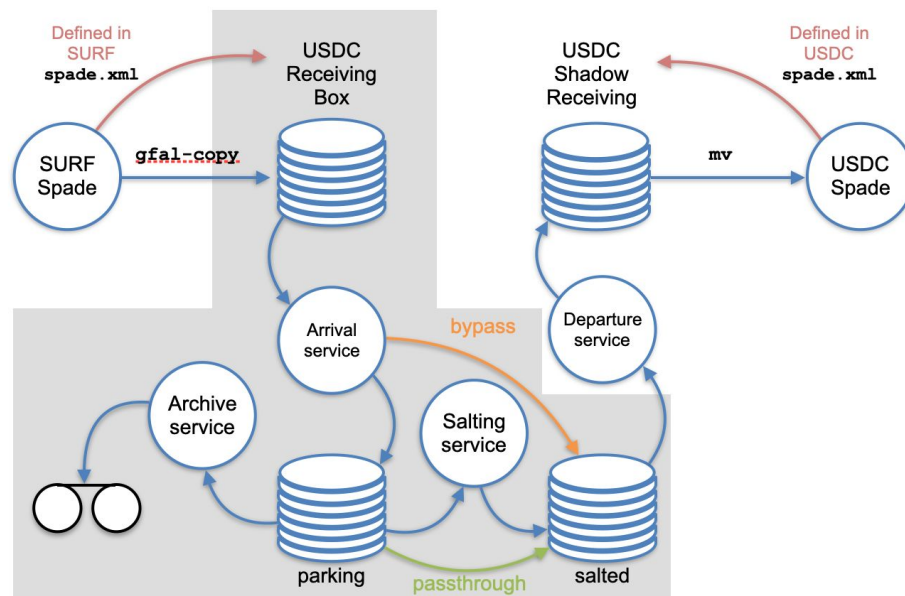
LZ Data Flow

- Fully Redundant design
 - **2 live copies of data** at US and UK data centers, 1 archived at US
 - Both data centers are able to perform data **simulation** and **(re)processing**, on demand
- **NERSC** and **GridPP** have diverging CPU architectures – all LZ software and analysis tools can run on either architecture
 - System choice is based on user preference



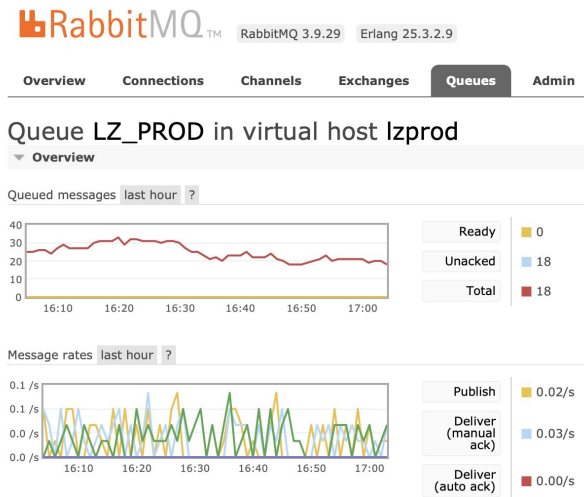
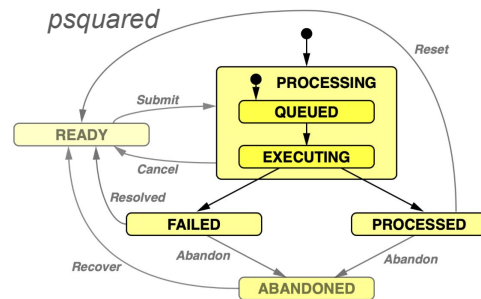
Data Movement at NERSC

- Automated data movement and bias mitigation service at NERSC
 - Java based custom application (SPADE) manages data movement and bookkeeping
 - SPADE has RESTful API interface
 - Salting service is our automated bias mitigation tool
 - Raw science data can be automatically salted before final placement and processing
 - Other types of data bypass salting service



Data Production at NERSC

- **Automated data processing** upon raw data arrival at NERSC
 - The workflow is managed using a **state machine** (*psquared* - custom software) with **RabbitMQ message queue**
 - Psquare has a RESTful API interface
 - Dedicated nodes on **Perlmutter at NERSC** for realtime job submission
 - Consumers managed by SLURM & RMQ
 - See *J.Siniscalco's poster this afternoon for sims prod at NERSC*
 - **Cataloging** of the raw and reconstructed data (mongoDB)
- **Web interfaces for monitoring** of data movement and processing
- This is accomplished by **extensive use of SPIN platform at NERSC**



Container-Base Science Gateway

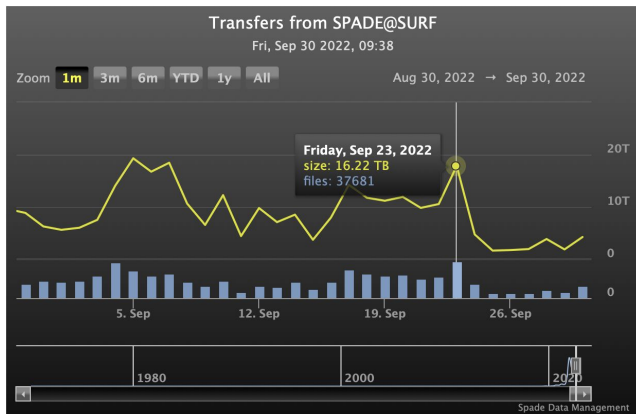
- LZ heavily relies on **SPIN** at NERSC (container-based services)
 - A container-based platform for **deploying science gateways**, **workflow managers**, databases, API endpoints, and web services
 - Based on Kubernetes, managed with the Rancher system
 - Services in Spin build with Docker containers
 - Data movement, processing, catalog, display, monitoring tools, databases



Namespace: lz-prod

<input type="checkbox"/>	Active	app-inspector	Deployment	registry.gitlab.com/nest.lbl.gov/images/app-inspector:3.16.0.v2	0	845 days	<input type="text"/>	⋮
<input type="checkbox"/>	Active	app-psquared	Deployment	registry.gitlab.com/luxzeplin/images/lzprod_lz-jee:1.0.0	0	265 days	<input type="text"/>	⋮
<input type="checkbox"/>	Active	app-read-only	Deployment	registry.gitlab.com/luxzeplin/images/lzprod_lz-jee:1.0.0	0	446 days	<input type="text"/>	⋮
<input type="checkbox"/>	Active	db-postgres	Deployment	postgres:14-alpine	0	967 days	<input type="text"/>	⋮
<input type="checkbox"/>	Active	db-rabbitmq	Deployment	rabbitmq:3.9-management	0	967 days	<input type="text"/>	⋮
<input type="checkbox"/>	Active	web	Deployment	registry.gitlab.com/nest.lbl.gov/images/nginx:1.22.0.v2-4096	0	966 days	<input type="text"/>	⋮

Monitoring of Data Movement and Processing



Run number(s) [Look up runs](#) [Look up runs](#)

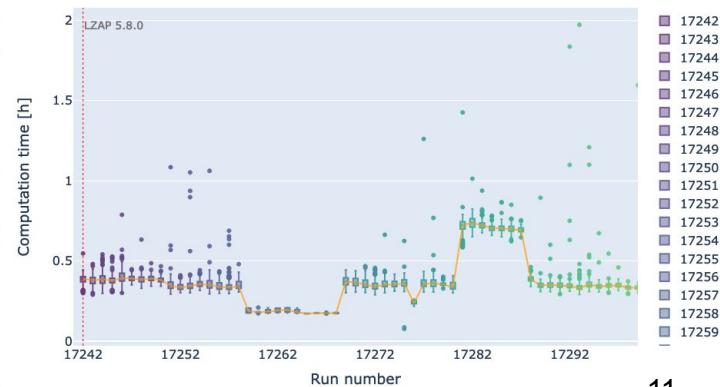
Bring runs missing files to top

Enter run numbers separated by spaces, then click the button at right

Run	Date and time of run start	SURF	USDC Placed	UKDC Placed
17408	2024/09/24 07:27:59 MDT	1200	1200	1190
17407	2024/09/24 03:53:33 MDT	1200	1200	1192
17406	2024/09/23 23:44:40 MDT	1200	1179	1179
17405	2024/09/23 23:35:29 MDT	40	40	40
17404	2024/09/23 23:34:12 MDT	40	40	40
17400	2024/09/23 23:03:29 MDT	40	40	40
17399	2024/09/23 22:54:39 MDT	40	40	40
17398	2024/09/23 22:45:39 MDT	40	40	40
17397	2024/09/23 22:37:30 MDT	40	40	40
17396	2024/09/23 22:28:05 MDT	40	40	40

Filter table by typing here:

Family	Start Date	Samples	Status	Progress
SR3_autodata_Y2024_W41_PROD-1	10-07-2024	15	○	2300 1 + expand
SR3_autodata_Y2024_W40_PROD-1	10-01-2024	71	○	11742 29 + expand
SR3_autodata_Y2024_W39_PROD-1	09-23-2024	72	○	10782 7 + expand
SR3_autodata_Y2024_W38_PROD-1	09-16-2024	63	○	6673 386 + expand
SR3_autodata_Y2024_W37_PROD-1	09-09-2024	64	○	11745 + expand
SR3_autodata_Y2024_W36_PROD-1	09-02-2024	54	○	11433 + expand



Data Visualization and Event Display

Very powerful and popular tool! Analyzers can type a run and event number for any collected data and access via web interface.

Event selection Pulse selection

Previous Event Next Event 7567 134816 Go

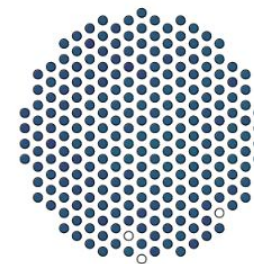
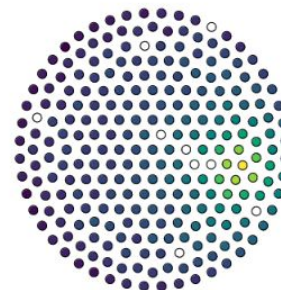
RQ file found in data catalog: /global/cfs/cdirs/lz/data/reconstructed/ WFs from RQ file

Run, Event ID: 7567, 134816	LZap 5.5.1 / condTag v968	Event 2317 / 2500 in RQ file
Global trigger	1646189658s, 349206780ns	2022/03/01 19:54:18 MT
Multiple Scatter	S1 pulse ID: 40	S2 pulse IDs: [45 47]
Weighted drift time: 8311.072 ns	S1c areas [phd]: [1517.2509 1511.911 S2c areas [phd]: [64688.63 285906.16]	

All TPC Skin OD Settings

Color Scale Min 57.342485427 Color Scale Max 33083.423854 Adaptive Scale

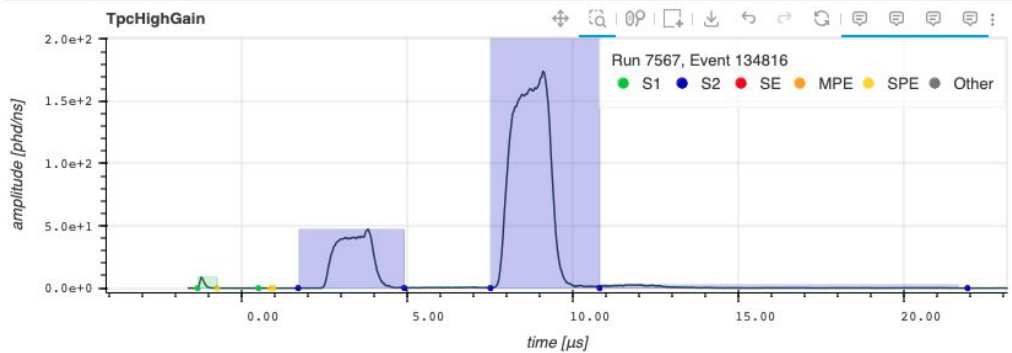
TPC PMT Arrays



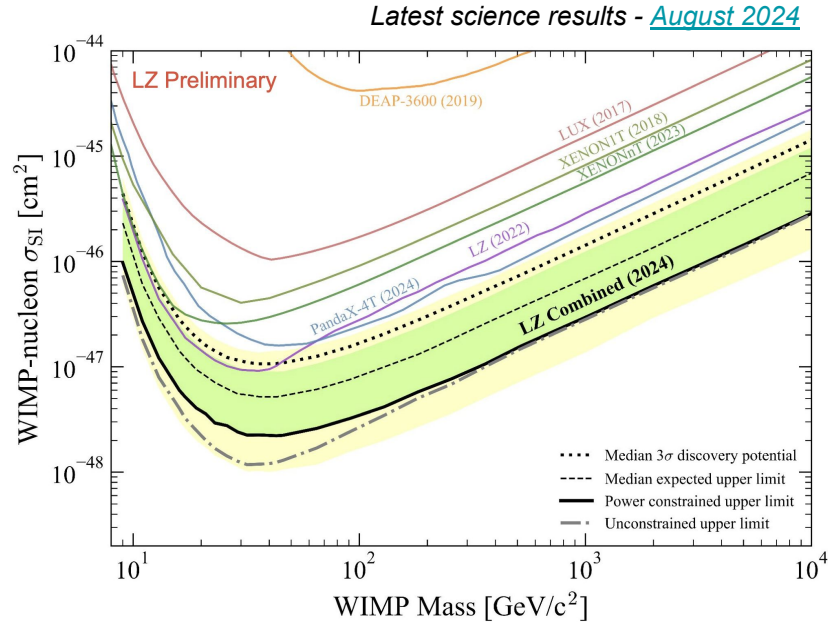
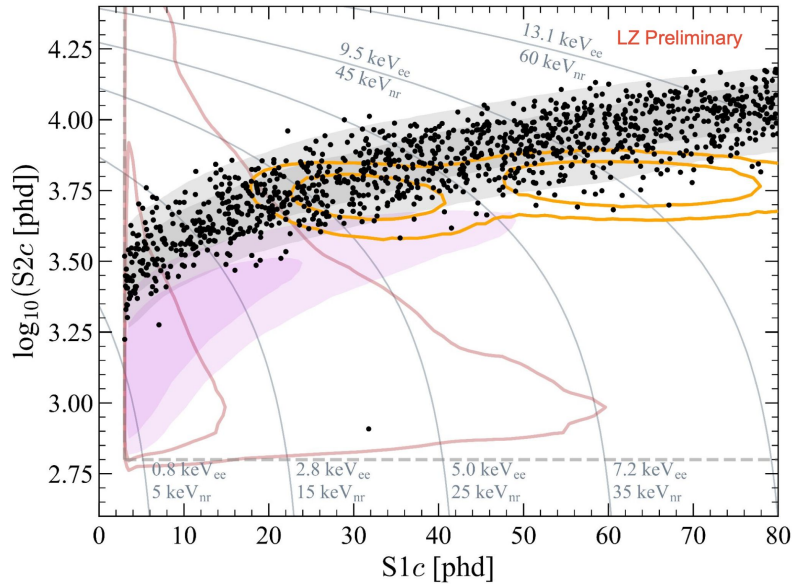
Summed Pods Individual Channels

Shade Pulses AFTs: 1 5 10 Hide buffer stop times Min [μs] Max [μs] Set x range

TPC HG TPC LG TPC HG JSON export



World-Leading WIMP Sensitivity



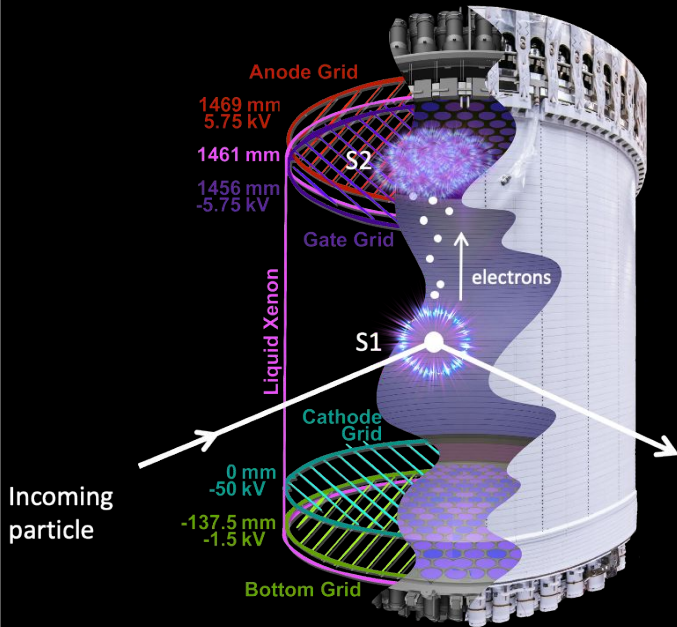
- 1220 events in WS2024 final data set
- 220 live days x 5.5 t = 3.3 tonne-yr

- Frequentist, 2-sided profile likelihood ratio
- Combined min cross section 2.2×10^{-48} cm² at 43 GeV/c²

Thank you!

LZ NERSC Operations Team: Maris Arthurs, Tyler Anderson, Keith Beattie, Ludovico Bianchi, Peter Gaemers, Steffen Luitz, Eli Mizrachi, Ibles Olcina, Simon Patton, Jacopo Siniscalco, and Maria Elena Monzani

Thanks to our sponsors and 38 participating institutions!



Incoming particle



U.S. Department of Energy
Office of Science



Science and
Technology
Facilities Council



Fundação
para a Ciência
e a Tecnologia



Swiss National
Science Foundation