Computing in Astroparticlephysics at MPI für Physik

Stefan Kluth
Oliver Schulz
MPI für Physik
NorduGrid 2018, 05.06.2018

Introduction
MAGIC/CTA
CRESST
GERDA/LEGEND
Conclusion

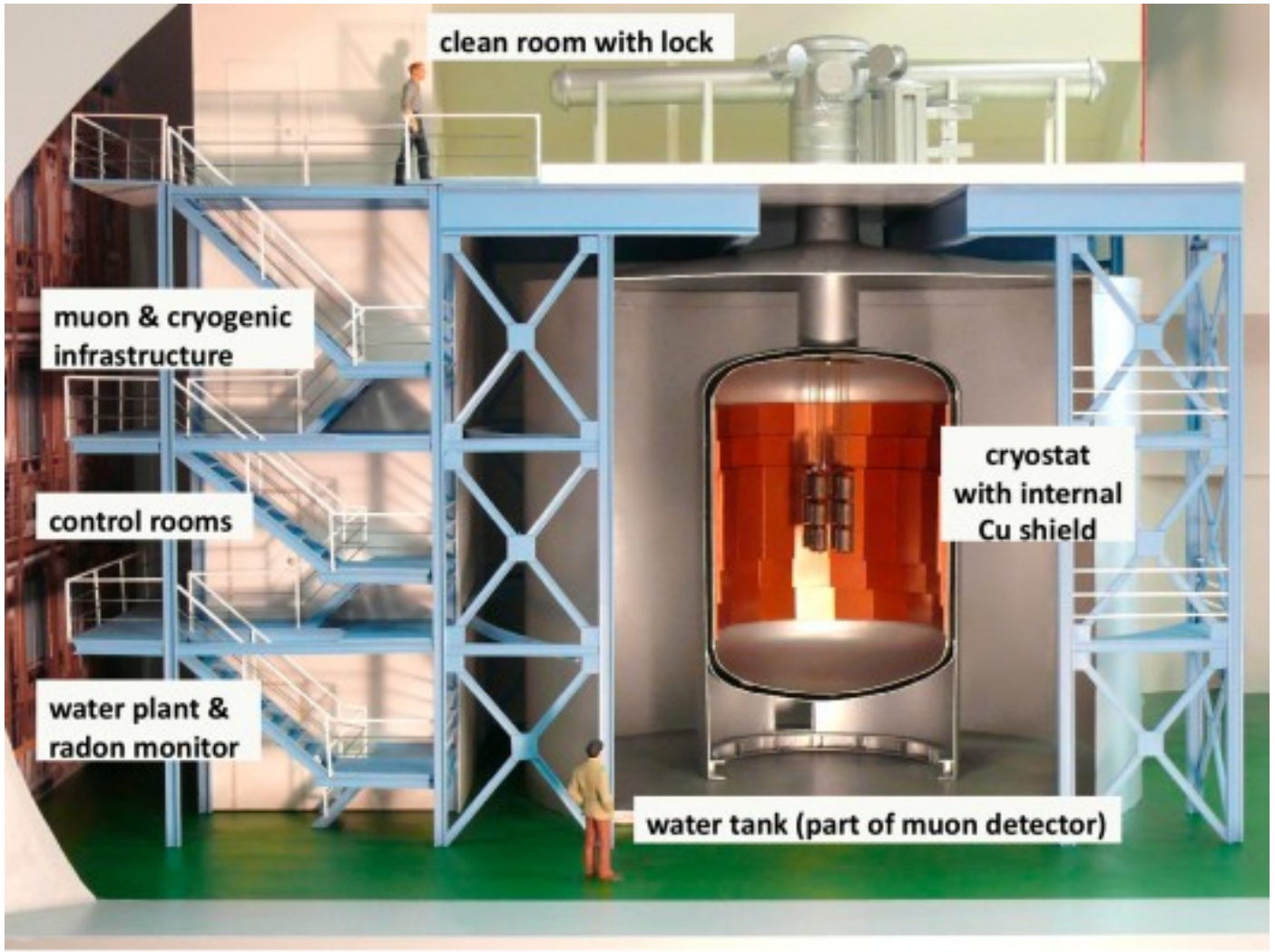
Introduction

- MPI für Physik
  - Department Bethke: ATLAS
  - Department Caldwell: GERDA, LEGEND, MADMAX, …
  - Department Teshima: MAGIC, CRESST, CTA, …
  - Theory departments: not (yet) computing intensive

GERDA

Search for neutrinoless double-β decay of Ge76

Majorana or Dirac neutrino → major impact on SMs of particle and astrophysics



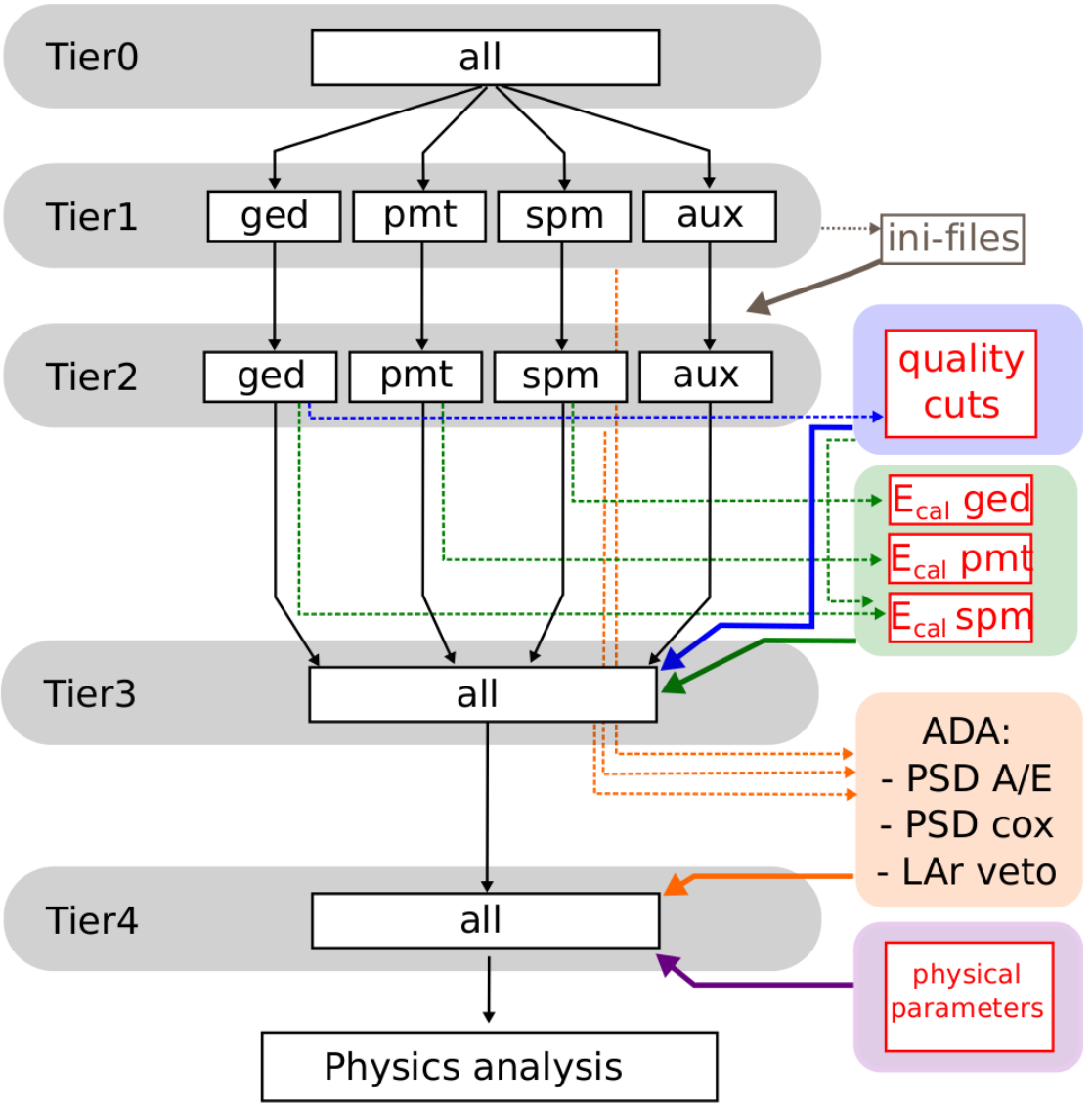
Ge76 based Semi-conductor Detectors

Triggered readout → comparatively low readout rate

Estimated 20 TB After completion

GERDA analysis workflow

TierN =
Data levels
like RAW,
ESD, AOD, etc



Metadata

GERDA analysis workflow

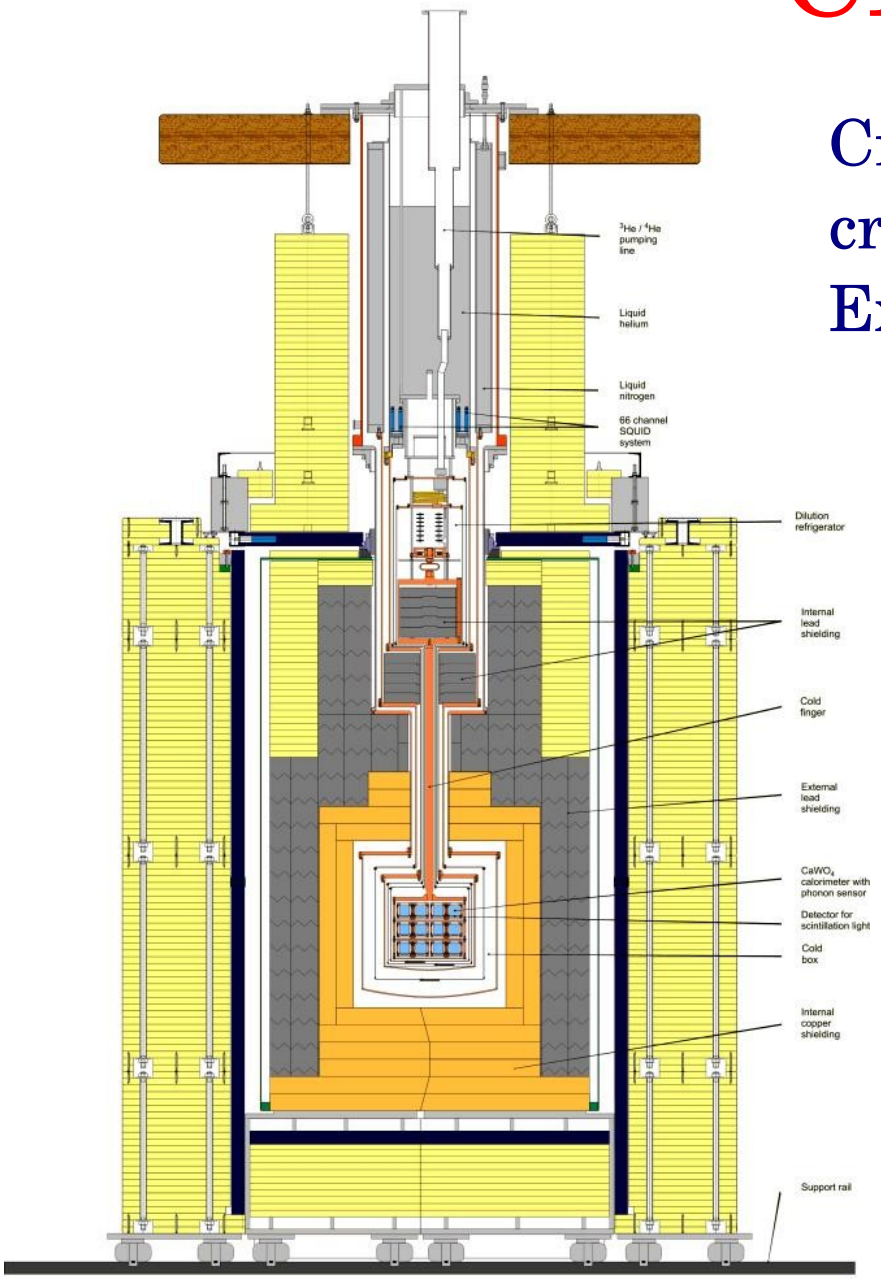
- Workflow orchestration with Luigi
  - Developed by Spotify
  - Individual steps are independent programs
  - Dependencies are DAG
- Luigi jobs can be sent to batch queue
- Filebased system, no need for DB services
- Conditions data not (yet) integrated
- Metadata (json) git version controlled
- No software framework

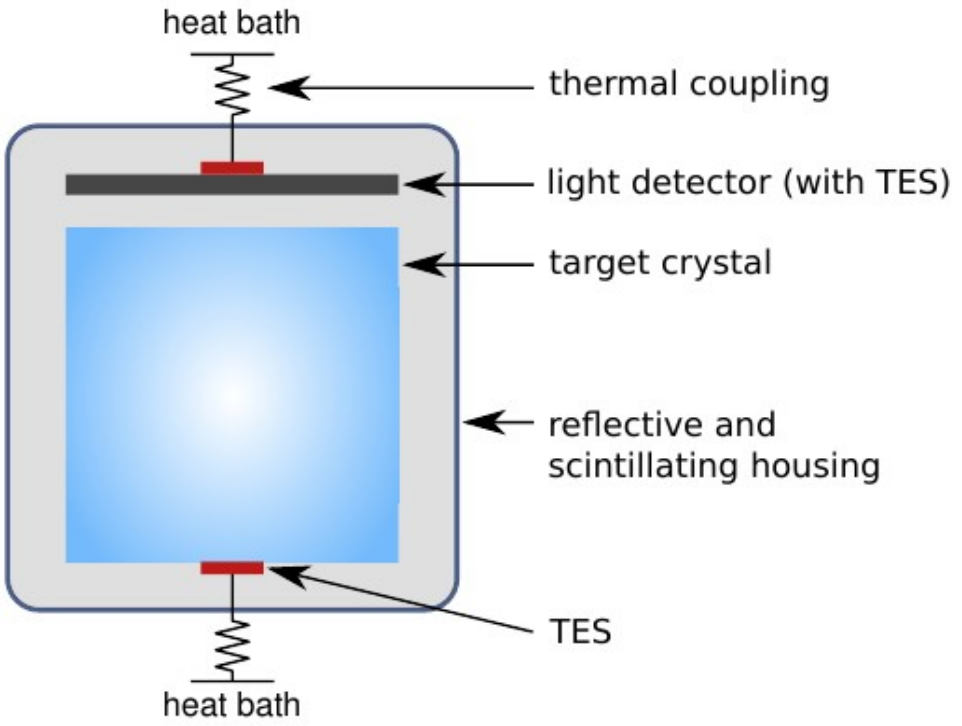
GERDA → LEGEND

- LEGEND
  - Much bigger successor to GERDA and similar
  - Expect O(1 PB) data for full-size LEGEND
  - Not yet detailed computing model
  - Some allocations at NERSC Cori for simulation
- GERDA computing managed at institutes
- MC production organised ad-hoc

CRESST

Cryogenic (lHe) detector with CaWO4 crystals to search for WIMP dark matter
Extremely low background





CRESST computing

- Experiment located at Gran Sasso lab
- Data transferred to MPCDF / MPP
- Recent move to untriggered FADC readout
- O(100 TB) per year, possibly more with larger detector
- Offline trigger must process all data before physics analysis
- Framework? Orchestration of components? Legacy software?

MAGIC

Cherenkov telescope for observation of high energy cosmic rays (La Palma island in the Atlantic ocean) 30 GeV to ~100 TeV

MAGIC

- Data from triggered readout processed on-site
- Data transferred to PIC (Barcelona)
- Data transferred from PIC to MPCDF (among others)
- At MPCDF
  - Several 100 TB data for analysis
  - Partially legacy software file based processing
  - Workflow orchestration by in-house scripting
  - Significant compute demand for air-shower simulation

CTA

- Much bigger successor to MAGIC and similar
  - 4 Large, several medium and many small telescopes
  - Expect WLCG like data volumes (~100 PB)
  - Common trigger of telescope array online
  - Dedicated WLCG-Tier2-sized computing container on site
  - Ship triggered/filtered/reduced data to CTA data centers (e.g. DESY)
  - Institutes? Software framework / workflow orchestration? Inclusion of medium and small telescopes for common events?

Software for small experiments

- Software development
  - Postdocs and PhDs
  - No (very few) full-time software&computing FTE
  - Not much automatic / formal testing of sw
- Software stacks
  - Small in comparison to LHC experiments
  - Possibly difficult to install
  - Some use of containers (docker, singularity) to manage software stacks

Ideal s&c for small experiments

- Personpower (FTE) limited
  - Cannot maintain dedicated IT infrastructure
- Cloud-like infrastructure (AWS, Google, …)
  - Data stored and managed simply and safely
  - Support file-based task oriented workflows
  - Easy export of data subsets to “laptop” for development
  - JupyterHub (e.g.) connection for interactive analysis
  - Services (github/lab etc.) connected

Does this exist?

- AWS could be used
  - S3 storage, workflow orchestration, web services (e.g. Jupyter) connection
  - Data management based on S3 API
  - Vendor lock-in?
  - After end of experiment (and funding) how to keep up AWS fees, or transfer to data preservation?
- CERN has all the components ...
  - EOS storage, xrootd tools, data management tools (e.g. rucio)
  - Lxplus/batch or openstack VMs for workflows
  - SWAN service

Does this exist?

- Unfortunately, most small experiments are not connected with CERN ...
- European (or national) science cloud initiatives?
  - If they will have a long life time
- Large computing centers?
- “Big” astro-particle labs?