Data Preparation for NA62

BR(\(K^+ \rightarrow \pi^+\nu\bar{\nu}\)) = (8.4 \pm 1.0) \times 10^{-11}
NA62 experiment: $K^+ \rightarrow \pi^+ \nu \bar{\nu}$

$\text{BR}(K^+ \rightarrow \pi^+ \nu \bar{\nu}) = (8.4 \pm 1.0) \times 10^{-11}$

- New in-flight decay technique
- Collect $10^{13}$ kaon decays, measure BR to 10%

- Precision kinematic reconstruction
- Precision PID
- upstream $K$
- downstream $e, \mu, \pi$

- GHz rates mean precise timing:
  - $\sim 100$ps between sub-detectors
- Hermetic $\gamma$ detection
- $10^7 \pi^0$ (from $K^+ \rightarrow \pi^+ \pi^0$) suppression
- $10^7 \mu$ suppression
Data reduction

K⁺→?  

<table>
<thead>
<tr>
<th>events / day (billions)</th>
<th>raw data / day (TB / day)</th>
<th>reco data / day (TB / day)</th>
<th>filtered reco / day (TB / day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>~1</td>
<td>10</td>
<td>20</td>
<td>10 (~ 10 * 1)</td>
</tr>
</tbody>
</table>

- Unit of data-taking is an SPS burst (spill), which lasts 3.5 seconds
  - 250k events per burst, take 4000 SPS bursts / day
  - Event size of ~10kB, ~2.5 GB per burst

- Fully reconstructed data in NA62 ~twice the size of raw (similar to ATLAS ESD), not kept

- Reconstructed data is filtered to reduce data volume by ~20, writing ~10 filters
  - ~200 physicists on NA62 start with filtered datasets
Data processing workflow

Key requirement:
- **100ps timing precision between sub-detectors**
- **Average t0s aligned at the 10ps level**

Workflow:
- Iterative calibration procedure performs coarse and then fine t0 calibration using 100 bursts, 1 DAQ run \( \approx = 1500 \) bursts
- Burst by burst calibration for high precision detectors during full event reconstruction to achieve 10ps alignment
- Post-production steps for filters and data quality
- Ad hoc steps/dependencies - robust production system

Baseline Conditions

Improved Conditions

<table>
<thead>
<tr>
<th>Raw</th>
<th>Reco</th>
<th>Filter</th>
<th>Hist</th>
</tr>
</thead>
</table>

DQ
ATLAS Tier-0 Production system

- Schematic of the original ATLAS Tier-0 PS design
- Over ten years of experience to be leveraged
- Core team very keen to collaborate (again!)
- …and close physical proximity at CERN helped
NA62 Production system

- NA62Transform python package to define workflows, turn inputs to outputs
- Custom TOM processes to orchestrate handshakes and auxiliary processes
- Running on custom VMs on CERN’s Openstack, including web servers
Production system performance

Web server uses shared puppet module between ATLAS and NA62

Production system can run at capacity since end of 2017

Final physics validation requires a lot of data, closing in on publication quality production now
Simplification

• Initial processing scripts were few thousand lines, starting with git checkout and compilation, sed for configuration changes, hard-coded if conditions, etc. (grew organically from best-effort work)

• Production system approach encourages separation of configuration and conditions, pre-built executables, i.e. conform to simple interfaces - *work that has been done by NA62 software experts*

• Current implementation has separated conditions handling into separate services, configuration is centrally managed

• Configuration of the production system (TOM and EOWYN) is via a templated file with only a handful of parameters, s/w release, conditions tag and processing attempt number
  • Also allow users to run exactly the same configuration within a sandboxed environment, useful for training and debugging

• Easy to use and reconfigure, (hopefully) also by non-experts
Production system configuration

Username - allow non-production roles
username '${TZUSERNAME}';

Declarative task definition - tie to username, control with pstates
inputdsspec `(username='${TZUSERNAME}' AND type='RAW' AND pstates LIKE '%calibso4:Done%')`;

Templated versioning and labelling for production system

tasktype ${TZUSERTAG}prod
tasktag p.${TZPROCVERSION}-${TZUSERTAG}${NA62SWRELEASE}

Transform job parameters, conditions, s/w release, job type and step

phconfig `{'phconfig': {"globalTag": "${NA62CDBTAG}", "release": "${NA62SWRELEASE}", "step": "Prod", "jobType": "Reco"}}`

- Username to allow non-expert production
- Declarative task declaration, SQL statement looking for matching input datasets, just in time task definition
  - pstates property used to control workflow (as well as production state for tasks)
- Minimal set of job parameters passed to NA62Transform code
  - NA62Transform takes input and produces output based on job type and step
HSF conditions data handling

From the HSF community white paper:
https://arxiv.org/abs/1712.06982

Database snapshots/dumps into filesystems or SQL files common request
Not least for disconnected running, HPCs

Data model of the schema lends itself well to a file-system layout

/\texttt{cvmfs/<experiment>/conditions/<gtag>/<system>/RunXYZ/payload}

NA62 is a running experiment

NA62 continues to use file-system storage for conditions, now using layout that maps to a database schema (allows future upgrade), with dedicated services to manage data access
Conclusion

• NA62 is a petabyte scale experiment with challenging data preparation requirements, *10ps cross-calibration of detectors*

• Collaborating with ATLAS colleagues, *a production system* has been implemented that manages a complicated workflow to process petabytes of data and *achieve the desired precision*

• *Separating conditions from configuration was key*, as was the centralisation of configuration of the software framework

• *Conditions data handling follows best-practice guidelines of HSF*

• Configuration of the production system uses only a handful of parameters and also allows users to run in a sand-boxed environment

• Hopefully this work ensures the successful processing of petabytes of NA62 data and helps pave the way to exciting physics results
Results

Standard Model theory prediction:

$$\text{BR}(K^+ \to \pi^+ \nu \bar{\nu}) = (8.4 \pm 1.0) \times 10^{-11}$$  [Buras et al. JHEP 1511 (2015) 33]

BNL measurement of kaons decaying at rest:

$$\text{BR}(K^+ \to \pi^+ \nu \bar{\nu}) = \left(17.3^{+11.5}_{-10.5}\right) \times 10^{-11}$$  [Phys. Rev. D 79, 092004 (2009)]

For comparison, interpret the event as a BR measurement:

$$\text{BR}(K^+ \to \pi^+ \nu \bar{\nu}) = 28^{+44}_{-23} \times 10^{-11} \ @ 68\% \ CL$$

Or given limited statistics of 2016 data, an upper limit:

$$\text{BR}(K^+ \to \pi^+ \nu \bar{\nu}) < 14 \times 10^{-10} \ @ 95\% \ CL$$

- Experimental technique works

Now only need to analyse petabytes of new data!