Data query and cleaning for Post Mortem beam loss analysis

Acknowledgments:
Christoph Wiesner
Daniel Wollmann
Michał Maciejewski
Philippe Bélanger

Fredrik Hülphers

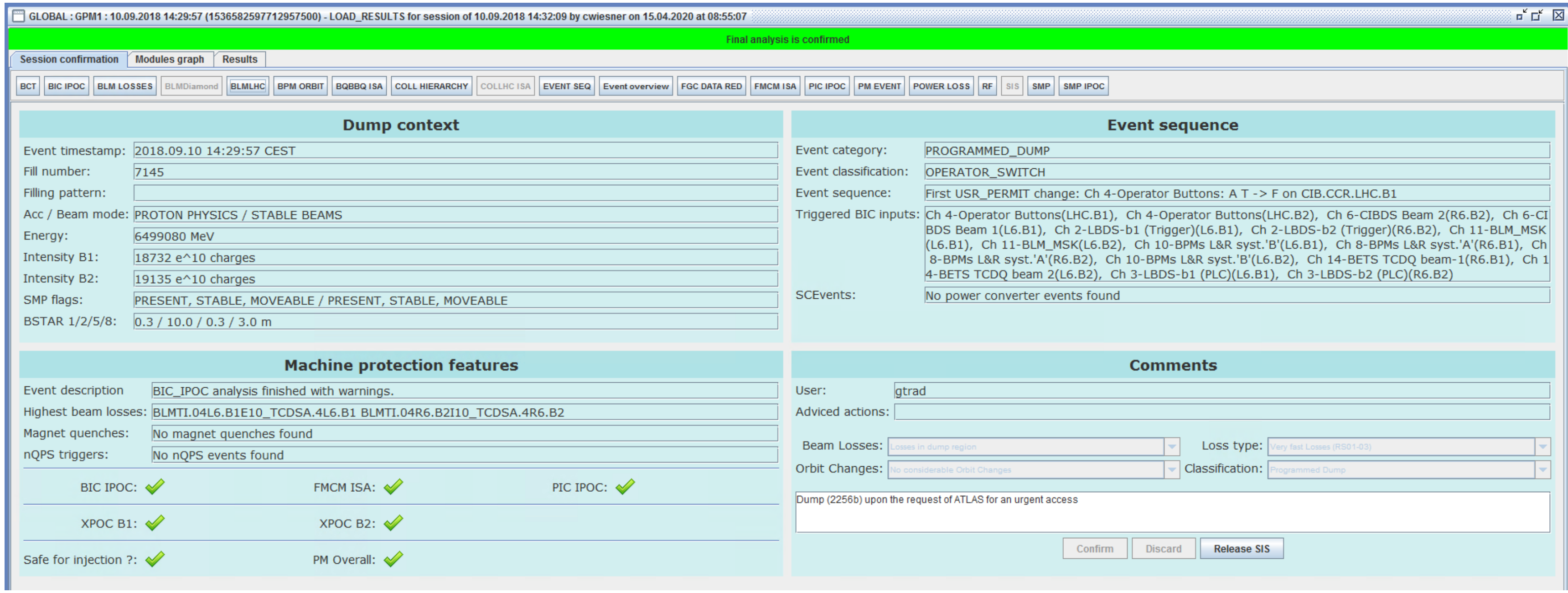
17/12/2020

Contents

Contents

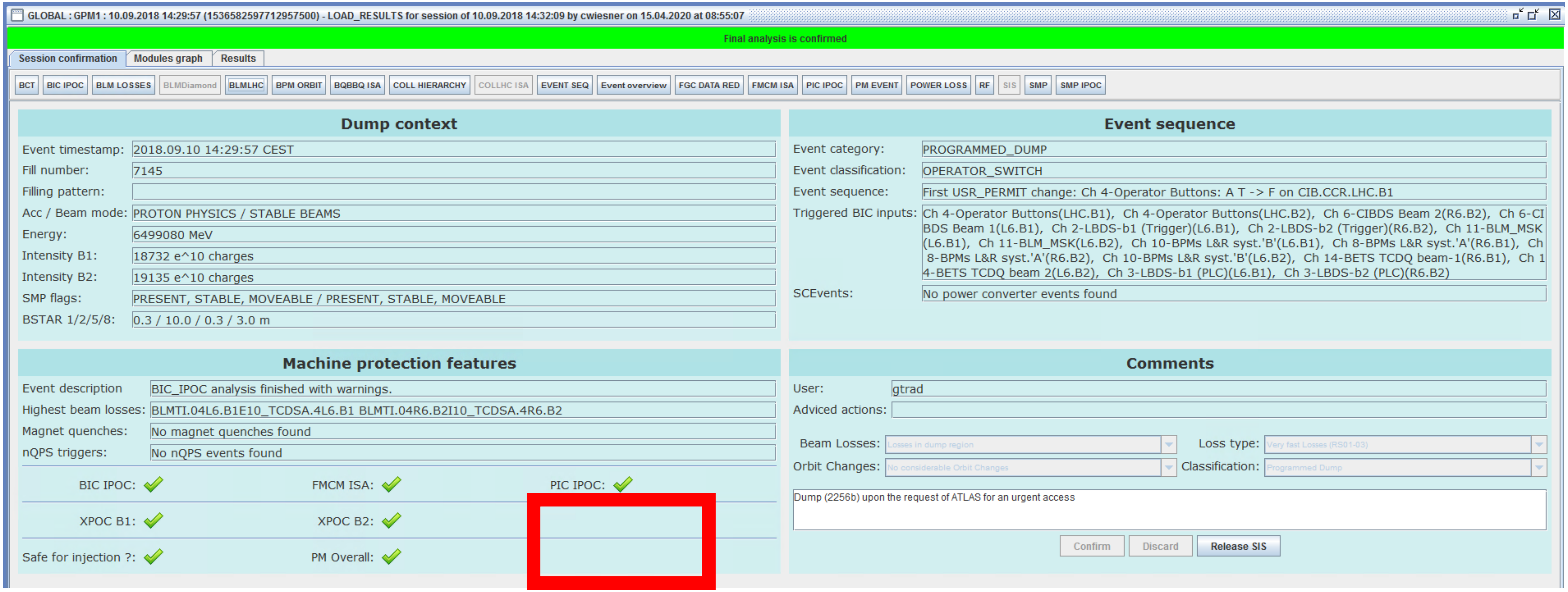
Introduction – What is Post Mortem?

- **Diagnostics system that stores data from *events*, most notably from LHC beam dumps**
- **Assert correct functioning of the Machine Protection Systems for all high-energy beam dumps, by machine protection experts**
  - Partially manual, partially automated
- **Identify potential issues**
- **PM Playback:**



Introduction – Project description

- **Goal: Develop an automated analysis tool for screening of beam loss data**



Introduction – Beam Loss Monitors (BLMs)

BLM type (in PM)	Count
IC	3651
LIC	110
SEM	177
Other	3230

3761

IC: Ionization chamber

LIC: Little ionization chamber

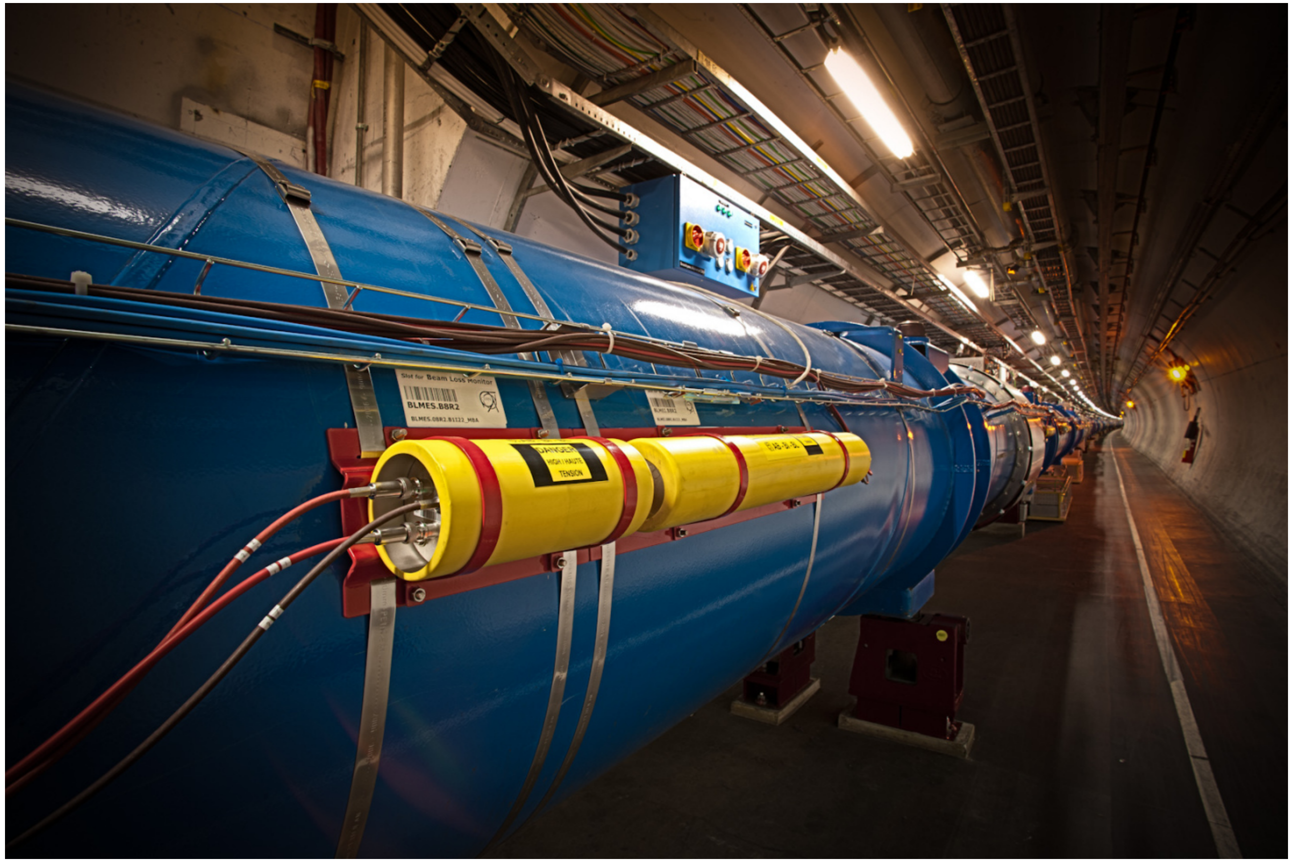
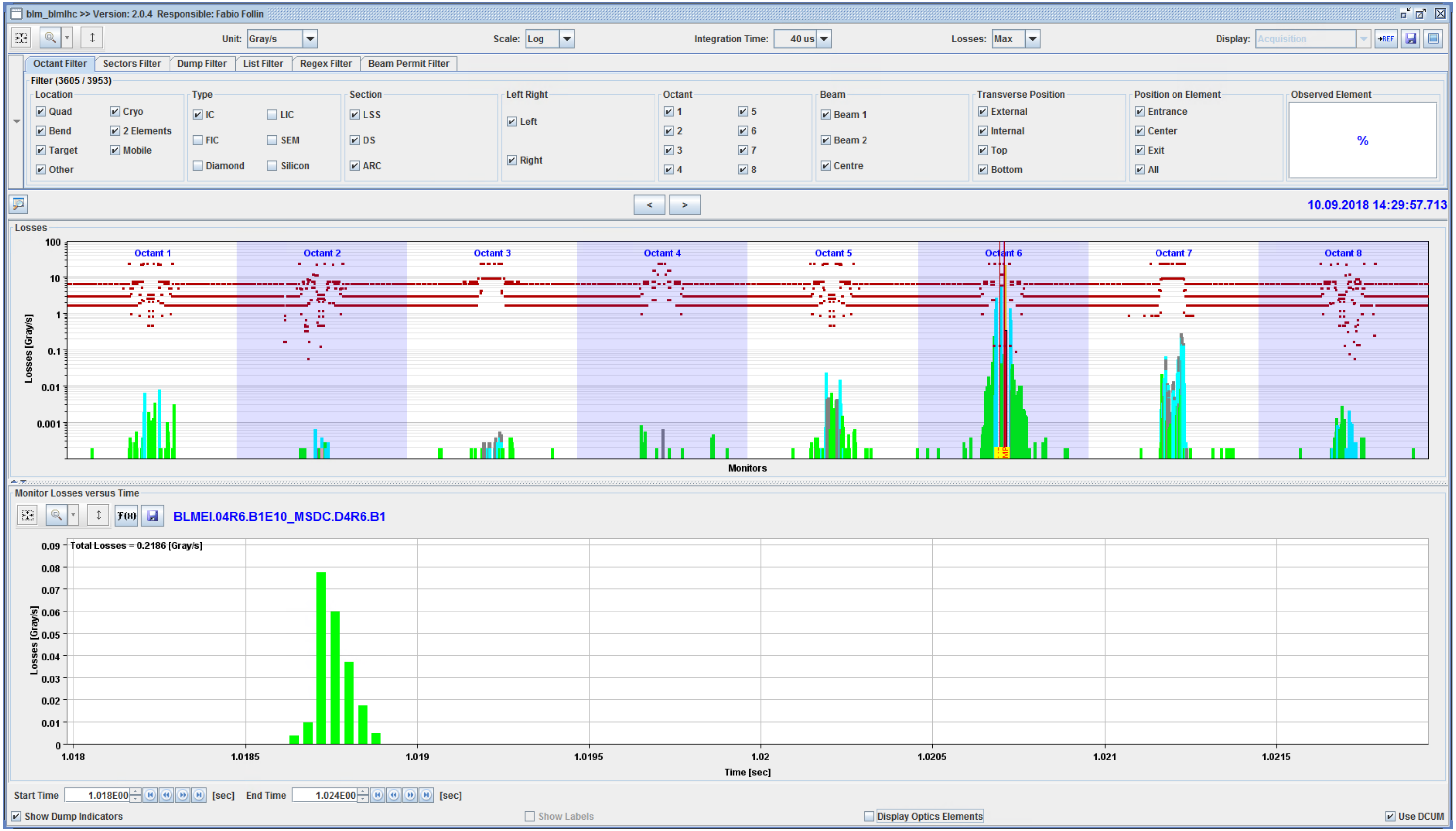


Image credit: Antonio Saba, CERN



blm_blmlhc >> Version: 2.0.4 Responsible: Fabio Follin
Unit: Gray/s
Scale: Log
Integration Time: 40 us
Losses: Max
Display: Acquisition
Octant Filter
Sectors Filter
Dump Filter
List Filter
Regex Filter
Beam Permit Filter
Filter (3605 / 3953)
Location
Quad
Cryo
Bend
2 Elements
Target
Mobile
Other
Type
IC
LIC
FIC
SEM
Diamond
Silicon
Section
LSS
DS
ARC
Left Right
Left
Right
Octant
1
2
3
4
5
6
7
8
Beam
Beam 1
Beam 2
Centre
Transverse Position
External
Internal
Top
Bottom
Position on Element
Entrance
Center
Exit
All
Observed Element
%
10.09.2018 14:29:57.713
Losses
Octant 1
Octant 2
Octant 3
Octant 4
Octant 5
Octant 6
Octant 7
Octant 8
Losses [Gray/s]
Monitors
Monitor Losses versus Time
BLMEI.04R6.B1E10_MSDC.D4R6.B1
Total Losses = 0.2186 [Gray/s]
Losses [Gray/s]
Time [sec]
Start Time
1.018E00
[sec]
End Time
1.024E00
[sec]
Show Dump Indicators
Show Labels
Display Optics Elements
Use DCUM

Introduction – Project outline

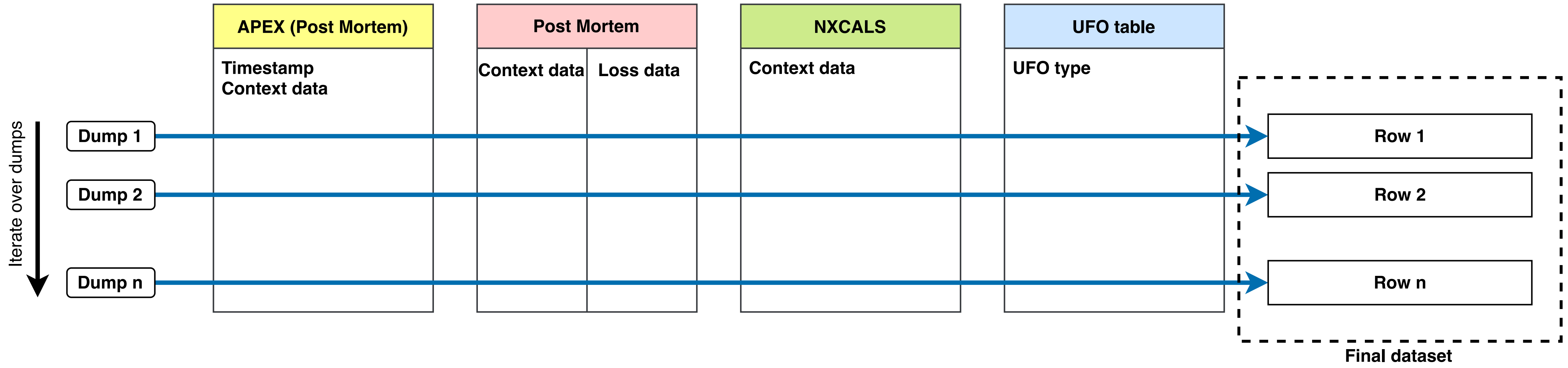
Steps:

- **Collect data from all beam dumps in Run 2 (2015 - 2018)**
  - Get **loss data** from BLMs
  - Get LHC **context variables**
    - E.g., beam energy, intensity, collimator positions
- **Use data to define "normal"/"acceptable" losses**
  - Approach 1: **Model-driven approach**
    - Based on understanding of physics and phenomena in the LHC
  - Approach 2: **Machine learning approach**
  - Probably combine the two approaches

Contents

Data collection

- **Data collection is done in several steps**
- **All data parts are merged in the end into one big dataset**
- **So far, I've only considered 2018 data**
    - Data from 2015-2017 to be included



Data collection - APEX

- **APEX: PM Database and Statistics - A tool for browsing PM data and statistics**
- **Get timestamps for all beam dumps**
- **20 context variables**
  - Beta star, operator comments, dump classification, …
- **Timestamps from APEX used as basis for further data collection**

	Event Timestamp	Event Category	Accelerator Mode	Beam Mode	Beam Energy [MeV]	Fill Number	Stable Beams [hours]	Fill Luminosity [nb^-1]	Mps Expert Comment	
	01-JUL-2018 09.46.24.305000	PROGRAMMED_DUMP	PROTON PHYSICS	STABLE BEAMS	6499200	6868	21.9	536.02875	Programmed dump of VdM fill with 140b. Clean dump.	525
	01-JUL-2018 11.56.34.773000	PROGRAMMED_DUMP	PROTON PHYSICS	FLAT TOP	6499200	6869	0	0	Asynchronous dump test at Flat Top 90 m run.	2nd
	01-JUL-2018 20.51.31.698927	PROTECTION_DUMP	PROTON PHYSICS	FLAT TOP	6499200	6871	0	0	Off-momentum loss map at Flat Top, nominal cycle, RF trim -500Hz.	2nd
	01-JUL-2018 22.16.01.889000	PROTECTION_DUMP	PROTON PHYSICS	INJECTION PHYSICS BEAM	449880	6872	0	0	-	25r
	01-JUL-2018		PROTON	INJECTION						

Post Mortem – Context data

- **Using a Python REST API to access PM data**
- **Script based on the LHC Signal Monitoring project**
  - Thanks, Michał!
- **SWAN notebook for launching script**
- **10 context variables**
  - Beam energy, beam intensity, abort gap population, …

	OVERALL_ENERGY	OVERALL_INTENSITY_1	OVERALL_INTENSITY_2	BEAM_MODE	timestamp_blm	aGXpocTotalIntensityB1	aGXpocTotalMaxInte
(7145,)	64990.8	18731	19135	11	1536582597713488525	9.132174e+07	3.9695

Post Mortem – Loss data

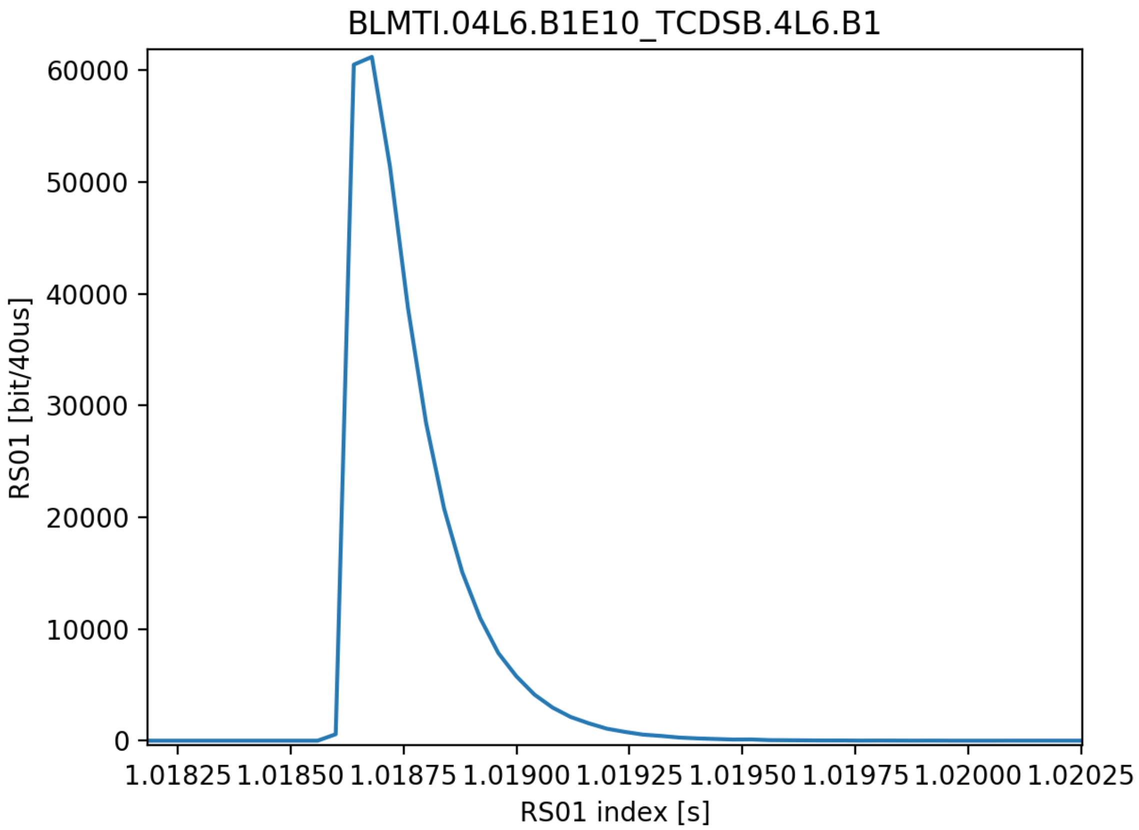
- **Losses recorded by each BLM are stored in PM in 2 running sums**

Running sum	Integration window	No. of values stored around dump	Length of time series around dump
RS01	40 µs	25600	1.024 s
RS09	1.31 s	512	671 s

- **RS01: 25600 values/BLM * 4000 BLMs * 1000 dumps/year ≈ 1e11 values/year**
- **Need to condense RS time series to make them more manageable**
- **Compute “features”**
  - Maximum, sum, rise time, fall time

Post Mortem – Loss data

- **Split RS01 time series in two (“before dump” and ”after dump”)**
  - Calculate features individually



NXCALS

- **A logging system using Hadoop Big Data technologies**
- **Continuously stores data during operation**
  - Context variables, BLM losses (stored at 1 Hz), …
- **Connect to Spark clusters via a SWAN notebook**
- **Using functions from the LHC SM API for querying the logging database**
- **Examples of context variables**
  - Collimator positions (LVDT gap) — Examples: TCTPH.4L1.B1, TCSG.D5L7.B2
  - Beam position (button BPMs) — Examples: LHC.BPTUH.A4L1.B1, LHC.BPTUV.D4R7.B2
  - Luminosity (at ALICE/ATLAS/CMS/LHCb)

Adding special flags - UFOs

- UFO: Unidentified Falling Object – Dust particle falling into the beam
- Can cause beam dumps
- Add three flags based on table compiled by Philippe Bélanger containing all UFO events in Run 2
- Get timestamp and UFO type (UFO, ULO, 16L2) from table

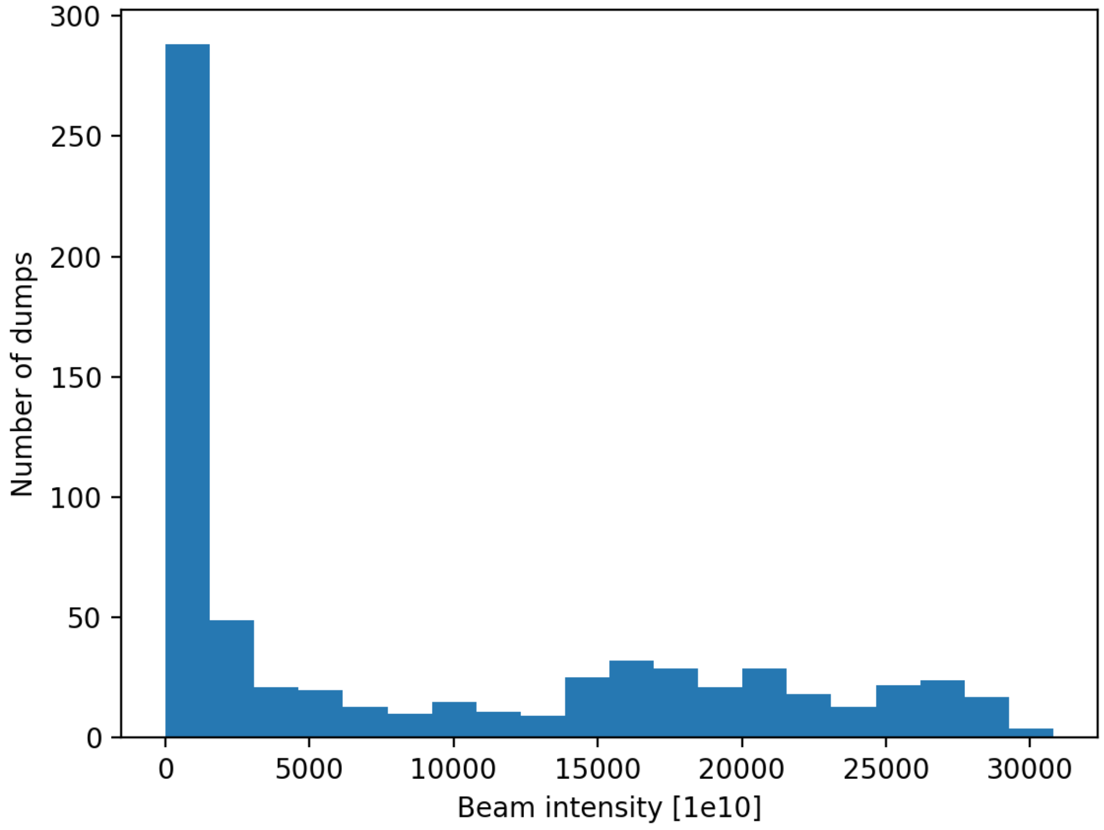
Timestamp	Type	Source/BLM	Comment
2018-07-17 22:07:21	16L2	HC.BLM.SR2.L/BLMQI.16L2.B1E30_MQ	Dumped by 16L2 losses when squeezing down, due...
2018-07-21 21:17:09	UFO	HC.BLM.SR8.C/BLMQI.01R8.B2E30_MQXA	Dumped by small UFO in IR8 and LHCb. Fast loss...
2018-08-07 18:31:17	16L2	HC.BLM.SR2.L/BLMQI.16L2.B1E30_MQ	16L2 B1 (2556b) @ 6.5TeV, ~5ms, dumped by TCP ...
2018-09-28 11:03:48	UFO	HC.BLM.SR7.R/BLMBI.18R7.B0T10_MBA-MBB_17R7	Beam dumped during squeeze, due to strong and ...

2018 dataset – Facts and figures

- Size of resulting dataset: **316 MB**
- 896 rows, 100 492 columns
- Estimated size without dimensionality reduction of loss data: **~500 GB per year**
- Number of UFO dumps: **21**
- Number of dumps with some missing PM context data: **140**

2018 dataset – Facts and figures

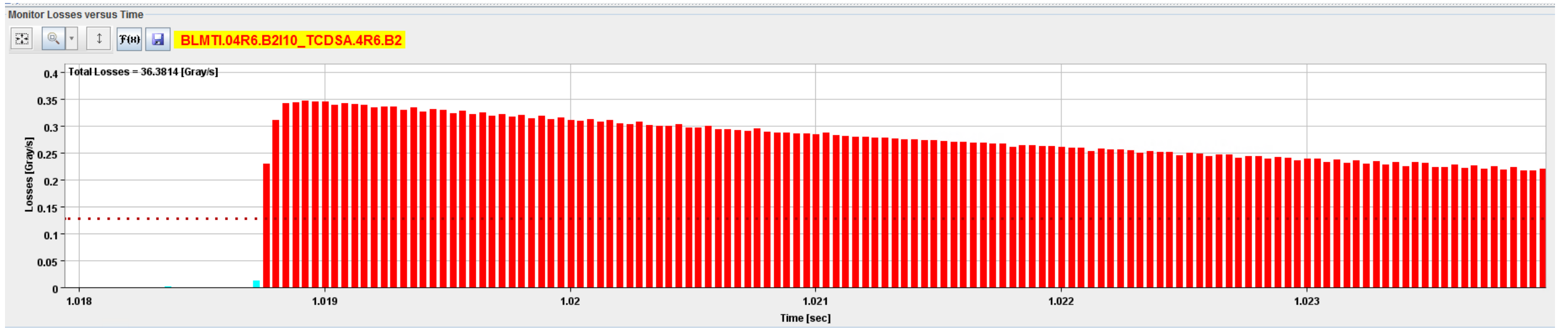
- **760 beam dumps in 2018 (when filtering out dump events without beam)**
- **By dump category:**
  - 406 programmed dumps
  - 354 protection dumps
- **By beam energy:**
  - 293 at injection energy (450 GeV)
  - 363 at top energy (6.5 TeV)
  - 104 between 450 GeV and 6.5 TeV
- **46 dumps had only one beam present**



Data collection – Dealing with issues

- **Handle missing data**
  - Missing context data from NXCALS and PM: Fill with N/A entries
- **Some context variables in NXCALS are stored with deviating frequencies**
  - Sometimes a variable doesn't have a recorded value close to the dump event
  - Sometimes the last stored value isn't accurate
- **Defining time index for "dump moment"**
- **Calculating rise and fall time for RS01**

	ALICE:LUMI_TOT_INST
-3.994	2.528513
-3.012	2.542154
-2.029	2.552235
-1.046	2.485550
0.046	1.208162
1.029	0.000000



Contents

Data cleaning – Initial checks

- **Ensure correct data types (integers, floats, strings, Booleans)**
- **Consistency checks:**
  - Beam intensity not above its theoretical maximum
  - Beam losses not above the BLM saturation level
  - Compare loss data from PM to loss data from NXCALS
    - PM: bits, 25600 Hz
    - NXCALS: Gy/s, 1 Hz
  - Manually evaluate RS01 max for each BLM over time to ensure that no BLMs were moved or had a filter installed/changed
- **Alignment checks:**
  - Make sure that data from the different sources are concatenated properly
  - Compare fill numbers and exact timestamps

Data cleaning

- **Filter out dumps where no beam was present**
  - Some non-dump PM events are stored as dump events
- **For each dump: Remove data from BLMs that had unphysical losses:**
  - Loss values above the BLM saturation level (165 BLMs in 2018)
  - Clearly abnormal loss values, e.g., only taking on values that are powers of 2
  - Replace unphysical values with N/A entries
- **Filter out BLMs that had no data at all in 2018**
  - 1640 BLMs (BLMM)

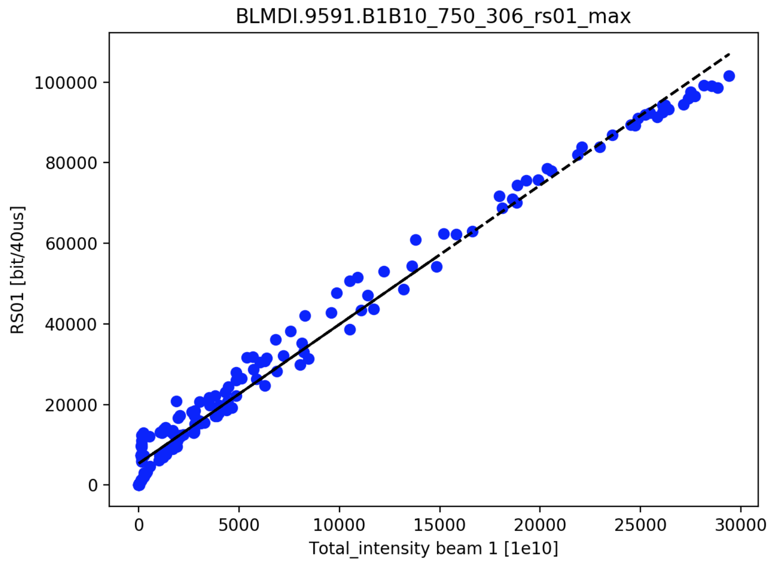
Contents

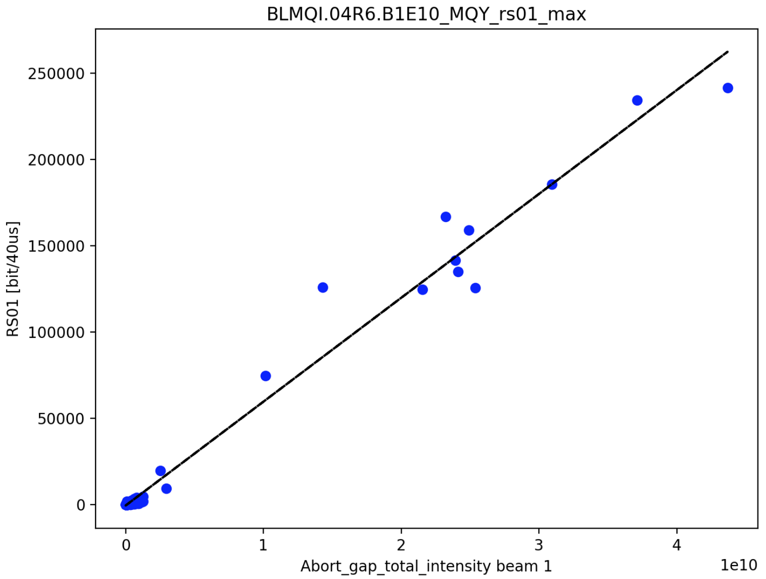
First look at the data – Filtering

- **Split dataset in two:**
  - Injection energy dumps (450 GeV)
  - High-energy dumps (>450 GeV)
- **Filter out BLM types other than IC and LIC**
- **For simple linear correlation checks:**
  - Filter out BLMs below noise level
- **Filter out dumps from the ion run**

First look at the data

- **Study first correlations between losses and some context variables:**
  - **Beam intensity:** 13 ICs strongly correlated (located in dump line or close to the dump)
  - **Abort gap population:** ~120 ICs strongly correlated





- **BLMs with consistently low losses (near noise level) are marked as "low-loss"**
  - 45 % of ICs and LICs had no RS01 losses greater than 25 bits (~2 mGy/s) in 2018

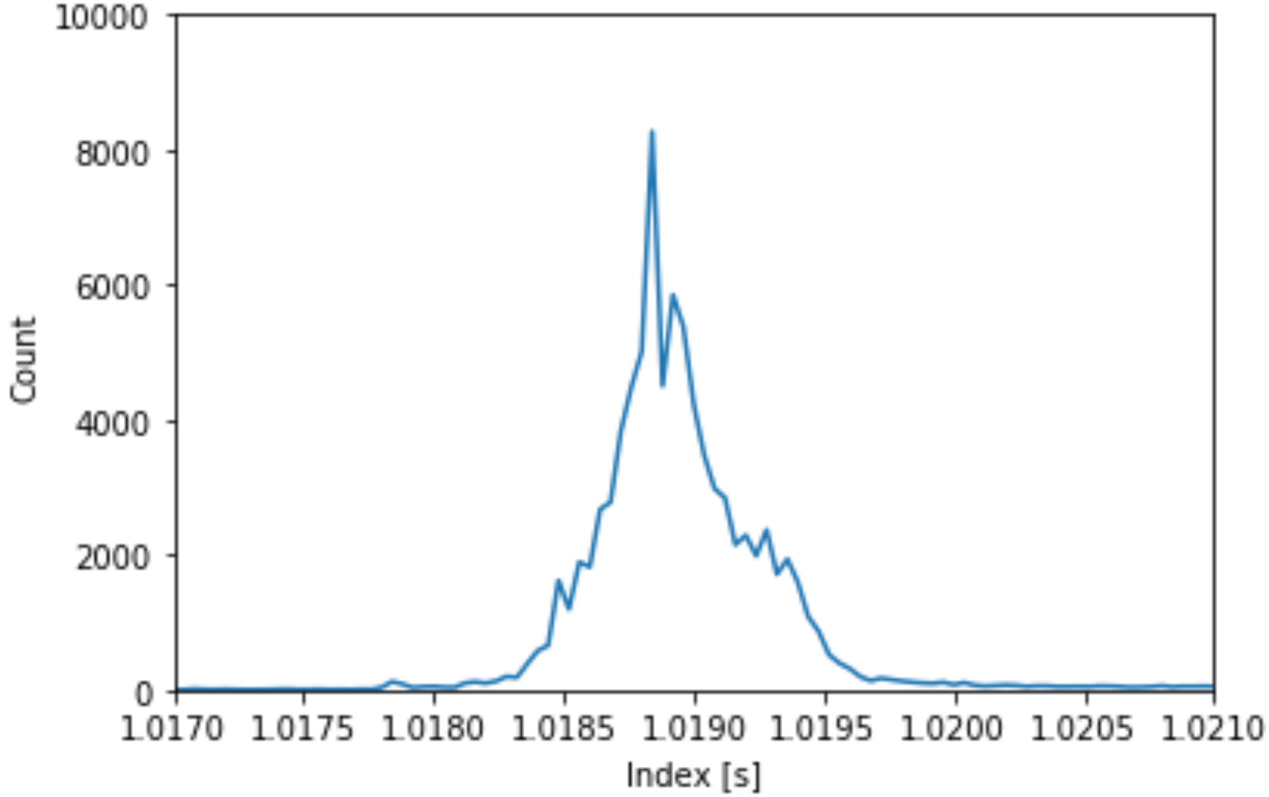
Contents

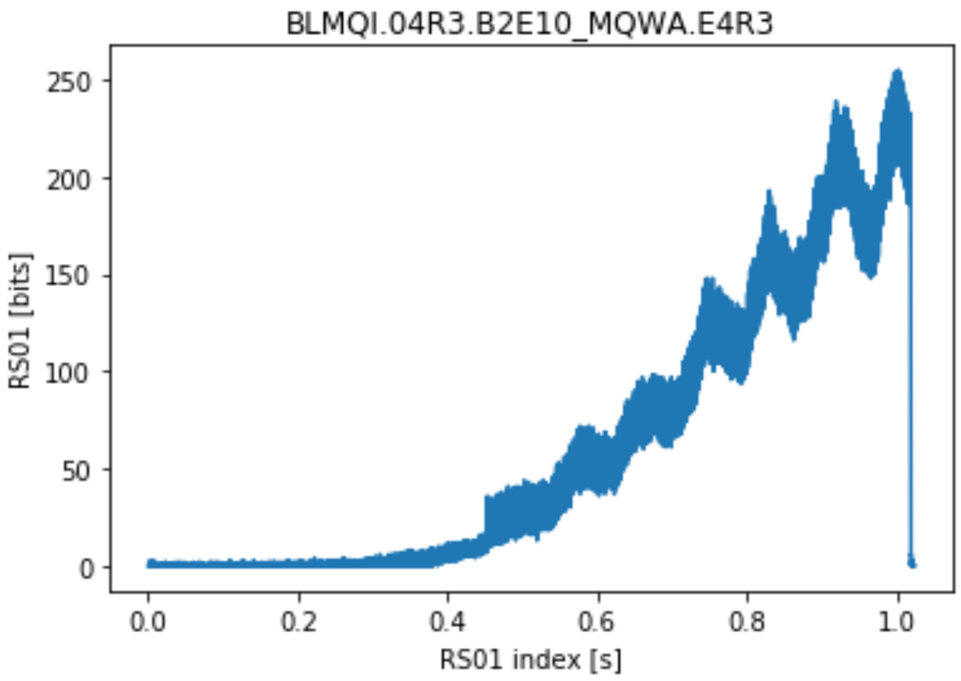
Outlook

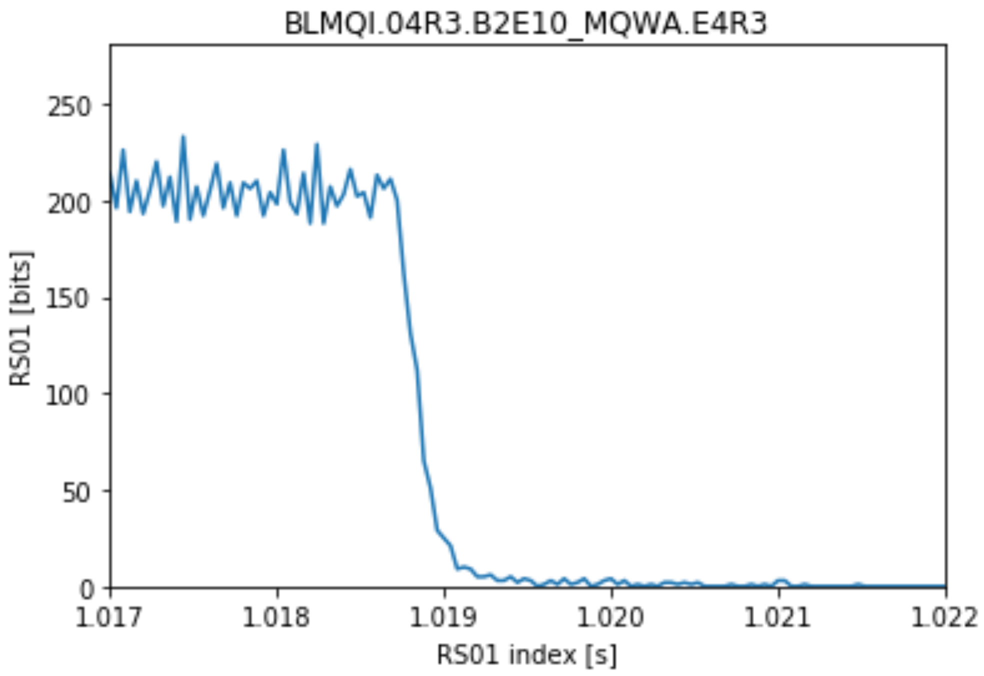
- **Collect data from 2015-2017**
- **Start defining simple thresholds for BLMs, based on collected data**
- **Perform some initial “predictions” and evaluate thresholds**
- **Post-processing of some vector-valued context variables**
  - Filling pattern, abort gap spill, bunch configuration, BPMs
- **In parallel to this: Run machine learning algorithms**
  - Neural networks
  - Anomaly detection, conditioned on the context variables
  - Distributed computing: HTCondor? OpenStack? Using GPUs if possible

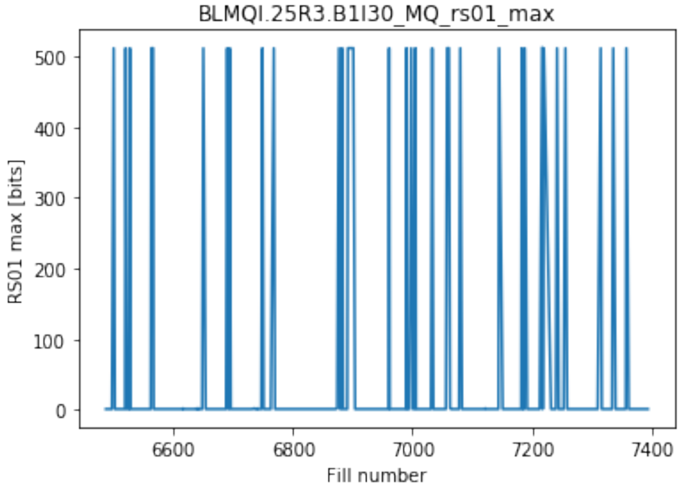
Thank you for your attention

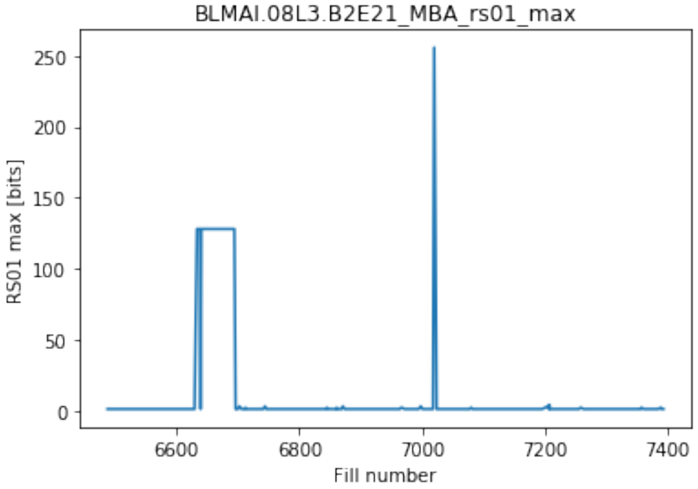
Questions?

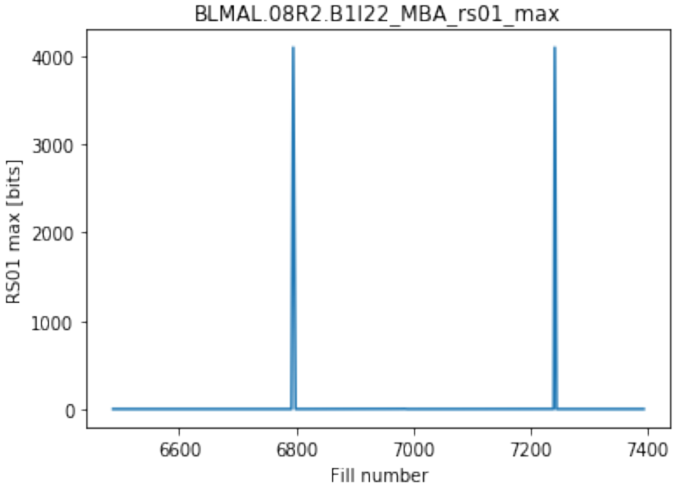

10000
8000
6000
4000
2000
0
Count
1.0170
1.0175
1.0180
1.0185
1.0190
1.0195
1.0200
1.0205
1.0210
Index [s]


BLMQI.04R3.B2E10_MQWA.E4R3
RS01 [bits]
RS01 index [s]
0
50
100
150
200
250
0.0
0.2
0.4
0.6
0.8
1.0


BLMQI.04R3.B2E10_MQWA.E4R3
RS01 [bits]
RS01 index [s]
0
50
100
150
200
250
1.017
1.018
1.019
1.020
1.021
1.022


BLMQI.25R3.B1I30_MQ_rs01_max
RS01 max [bits]
Fill number


BLMAI.08L3.B2E21_MBA_rs01_max
RS01 max [bits]
Fill number


BLMAL.08R2.B1I22_MBA_rs01_max
RS01 max [bits]
Fill number