

# A data analysis competition

Nikita Kazeev Alexey Rogozhnikov Andrey Ustyuzhanin

The Yandex School of Data Analysis

August 27, 2015

- ▶ A planned experiment at J-PARC, Japan
- ▶ Searches for coherent neutrino-less conversion of a muon to an electron,  $\mu^- + N(A, Z) \rightarrow e^- + N(A, Z)$ .
- ▶ The process breaks the leptonic number conservation law.
- ▶ Previous upper limit was set by SINDRUM II in 2006, COMET is going to have 10000 times better sensitivity.

# Experiment idea

Smash muons into aluminum to make muonic atoms, see if any electrons fly out.

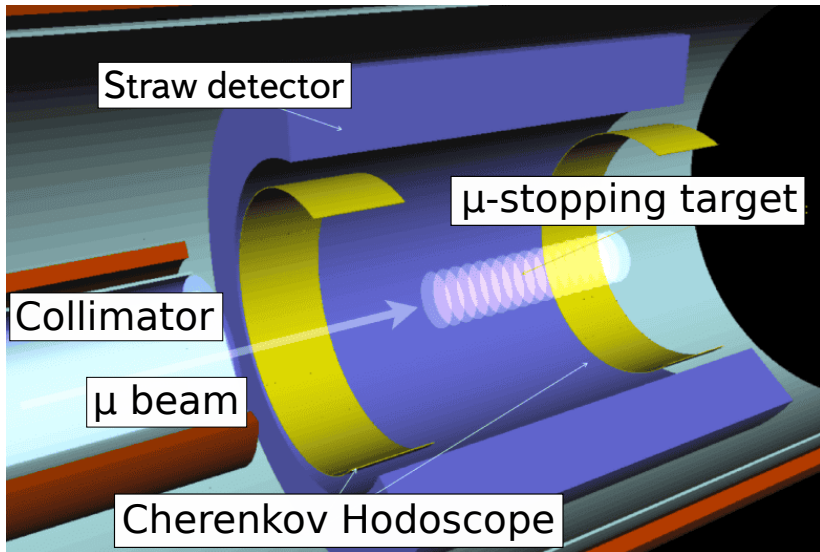
Standard model, peak at 52.8 MeV:

$$\mu^- \rightarrow \nu_\mu + e^- + \bar{\nu}_e \quad (1)$$

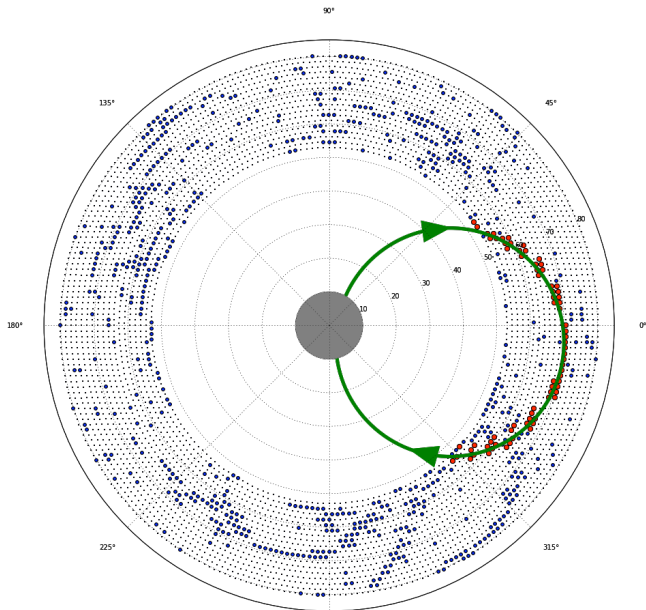
Charge lepton flavour violation, peak at 105 MeV:

$$\mu^- + \text{Al} \rightarrow e^- + \text{Al} \quad (2)$$

# Detector



# Event



- ▶ A Monte-Carlo simulation
- ▶ An event is a 'snapshot' of detector, it consists of data taken from all 4482 wires
- ▶ "energy" - energy deposited at the wire in GeV
- ▶ "relative\_time" - difference between time the straw tube detected energy and time of hodoscope in ns
- ▶ "label" - 0 for inactive wires, 1 for hits that are a part of a track, 2 for background hits

## Competition - basics

- ▶ You are given some labeled data - train dataset
- ▶ You are given some unlabeled data - test dataset
- ▶ You are to predict the probability of being signal for hits from the test dataset
- ▶ You can work in team of two
- ▶ You are to receive a prize if you make into top-3 of your track
- ▶ You are to present your solution at the award ceremony - if you win

## Competition - technicalities

- ▶ The link is in the StarterKit and the school website
- ▶ The test dataset is divided into two parts: public and private
- ▶ On the leaderboard you can view the score you get on the public part
- ▶ The final standings will be calculated on the private part
- ▶ The public/private split is based on individual hits, not events
- ▶ Private code sharing is not allowed - if you share, you must share with everybody via the forum



# Acknowledgments

We thank COMET collaboration (and specially Chen Wu) for allowing us to use this dataset, Ewen Gillies (ICL) for collaborating on the analysis it and Kaggle Inc. for hosting the competition.