# Parallel data processing on the Prometheus Supercomputer to aid proton beam studies in single particle detection mode using the scCVD Diamond Detector

Kinga M. Jeleń[1], Filip Dej[2], Marcin Jastrząb[1], Jan K. Kapała[2], Tomasz Nowak[1], Paweł Olko[1],

Marzena Rydygier[1], Mateusz B. Szarek[2], Mateusz Szpyrka[2], Patryk Wójtowicz[2],

Michael P.R. Waligórski[1], and Leszek Grzanka[1]

[1] Institute of Nuclear Physics Polish Academy of Sciences, Krakow, Poland
[2] student at AGH University of Science and Technology, Krakow, Poland
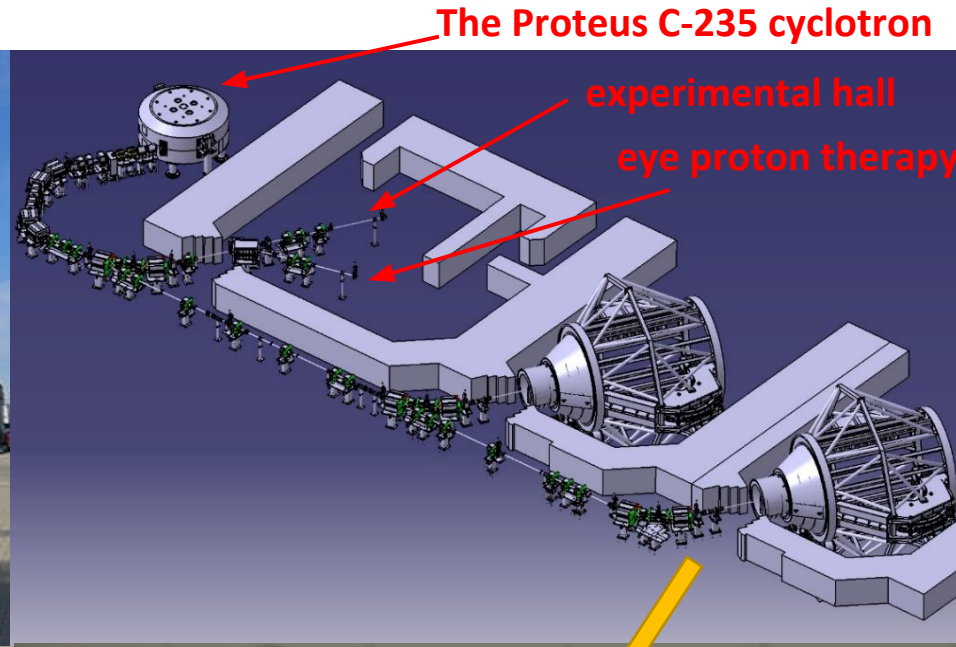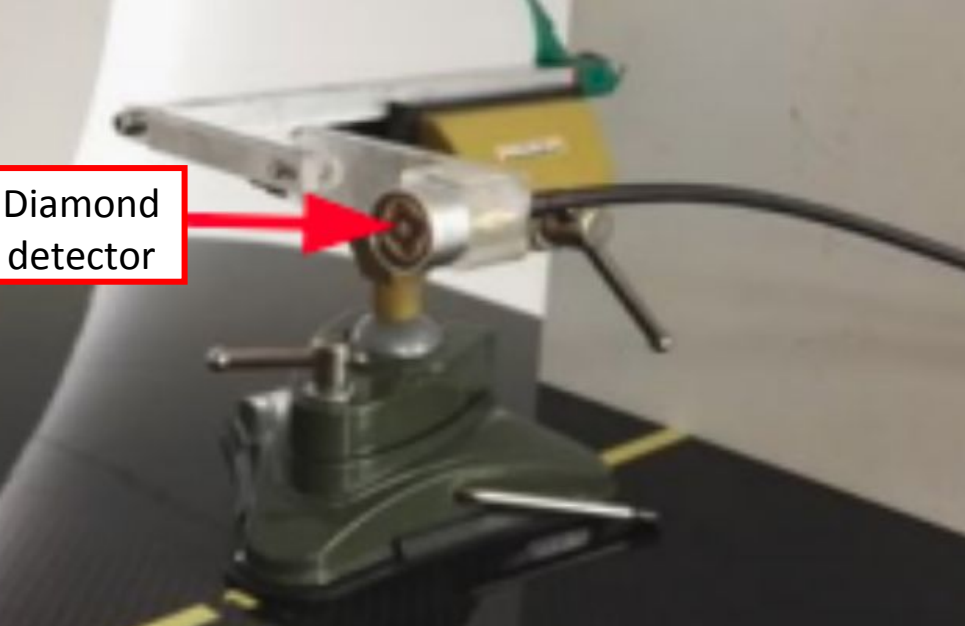
email: kinga.jelen@ifj.edu.pl
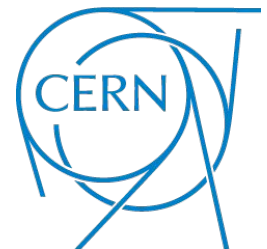
KUKDM 2018, Zakopane

Proton radiotherapy at IFJ PAN

The Proteus C-235 cyclotron

experimental hall

eye proton therapy

Bronowice Cyclotron Center

Gantry
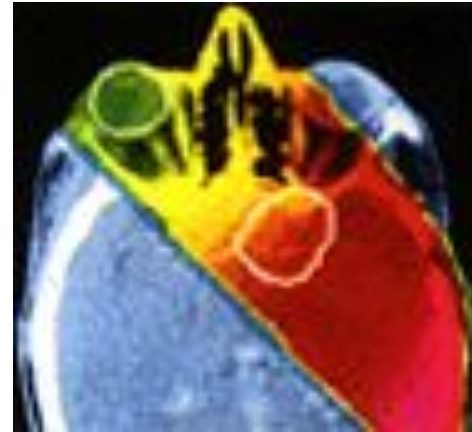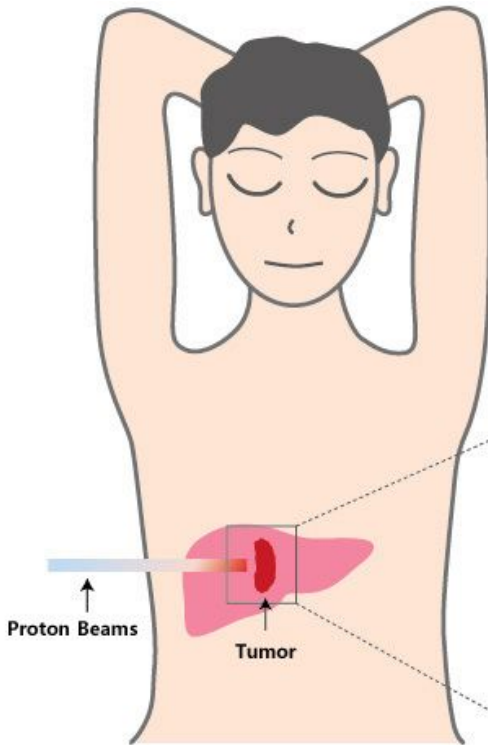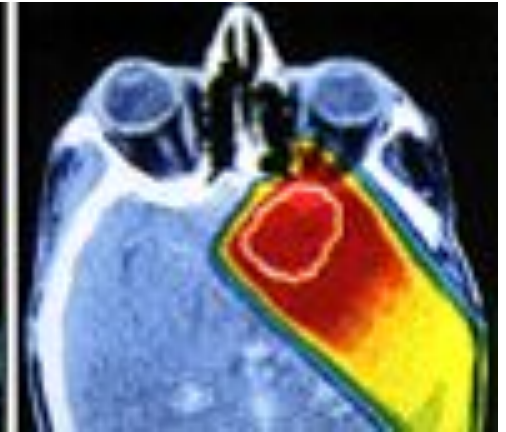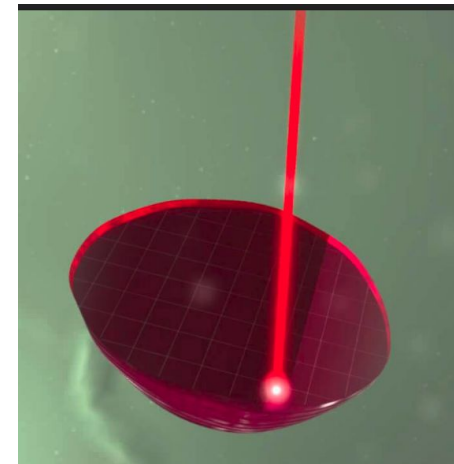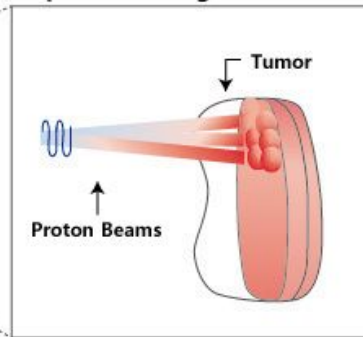
Diamond detector

Dose distribution in radiotherapy

gamma-ray beam

proton beam
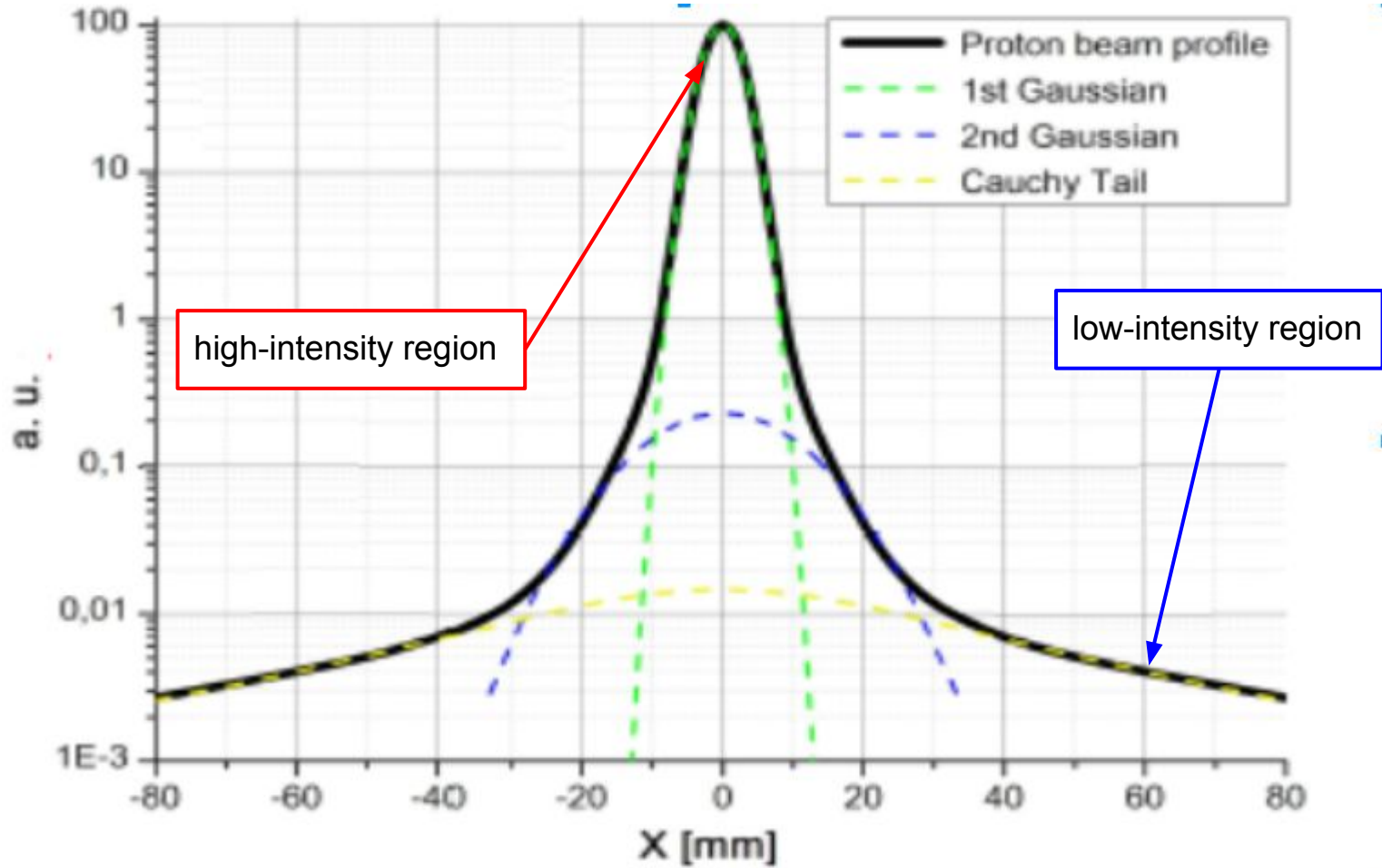
Spot-Scanning Irradiation

Tumor

Proton Beams

Proton Beams

Tumor

Stopping protons can save organs at risk!

Measurements of the narrow proton beam

It is necessary to obtain a very precise information about **low-intensity spot regions** of Proton Beam Scanning.

Diamond detector wide dynamic range

**WHY scCVD Diamond
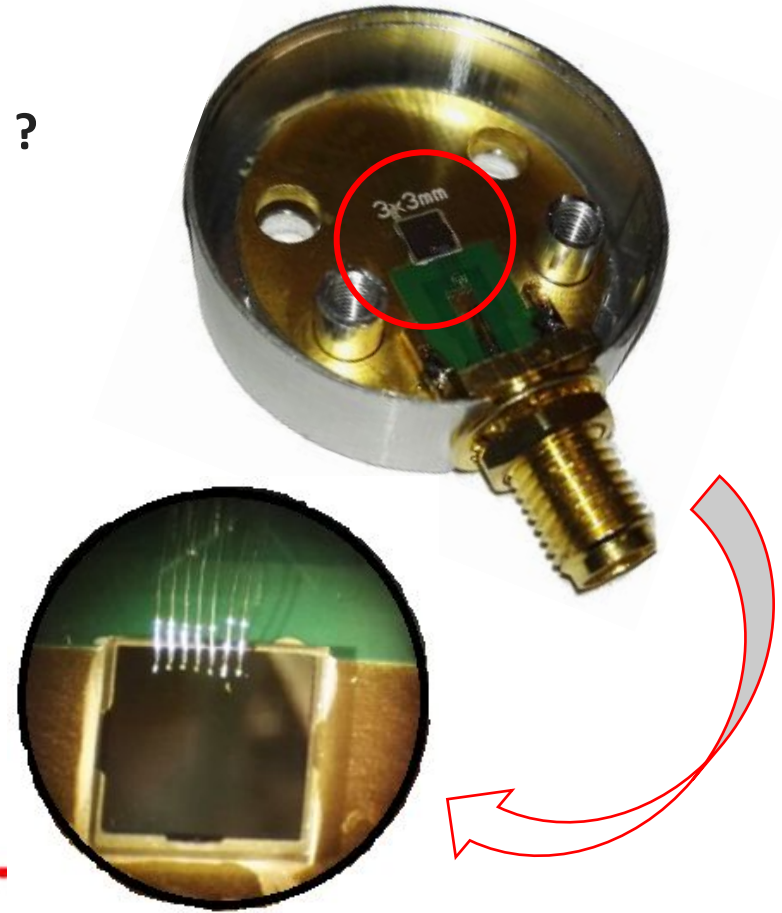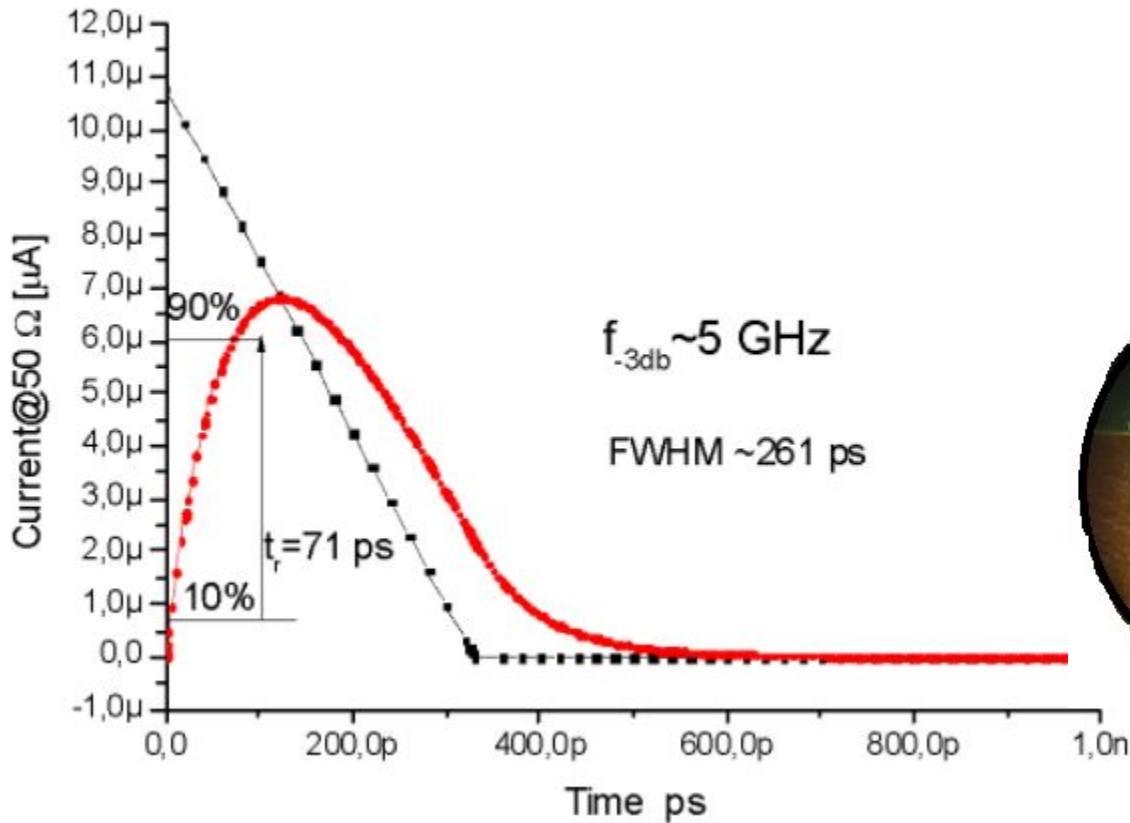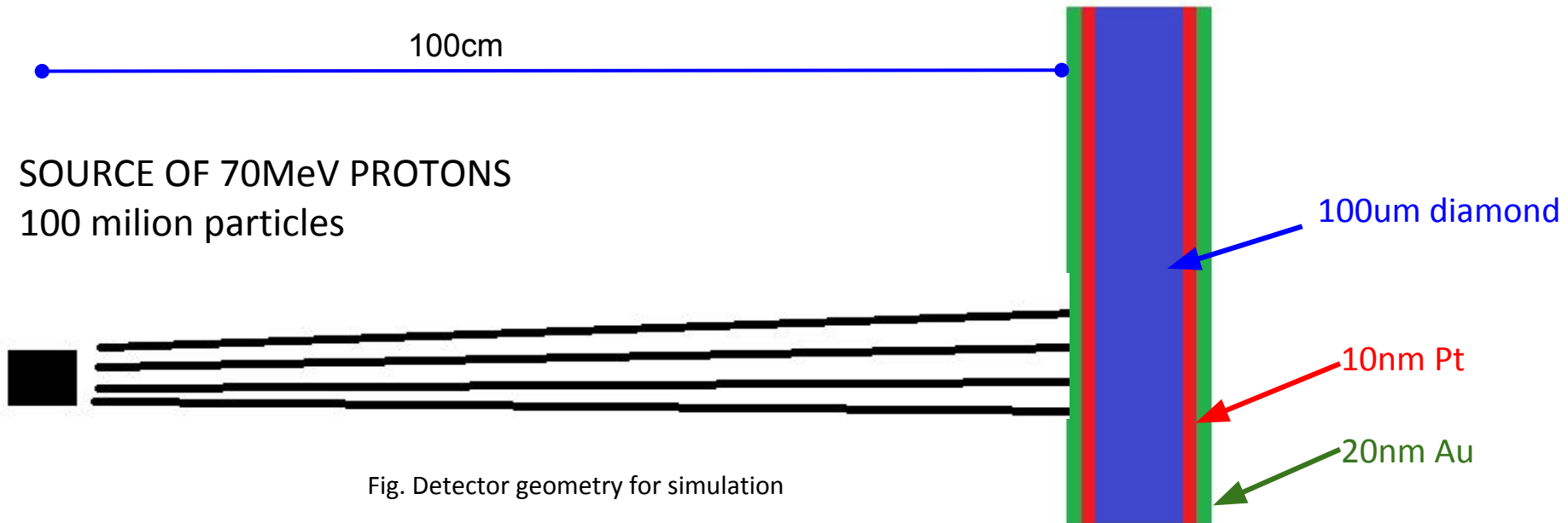(single crystal Chemical Vapour Deposition) ?**

$f_{-3db} \sim 5$ GHz

FWHM ~ 261 ps

$t_r = 71$ ps

Fig. Diamond detector package and its sensitive area

It generates 5 GB/s. This is why we need Supercomputer!

Simulation of diamond detector response using MC SHIELDHIT-12A

100cm

SOURCE OF 70MeV PROTONS
100 milion particles

100um diamond

10nm Pt

20nm Au

Fig. Detector geometry for simulation
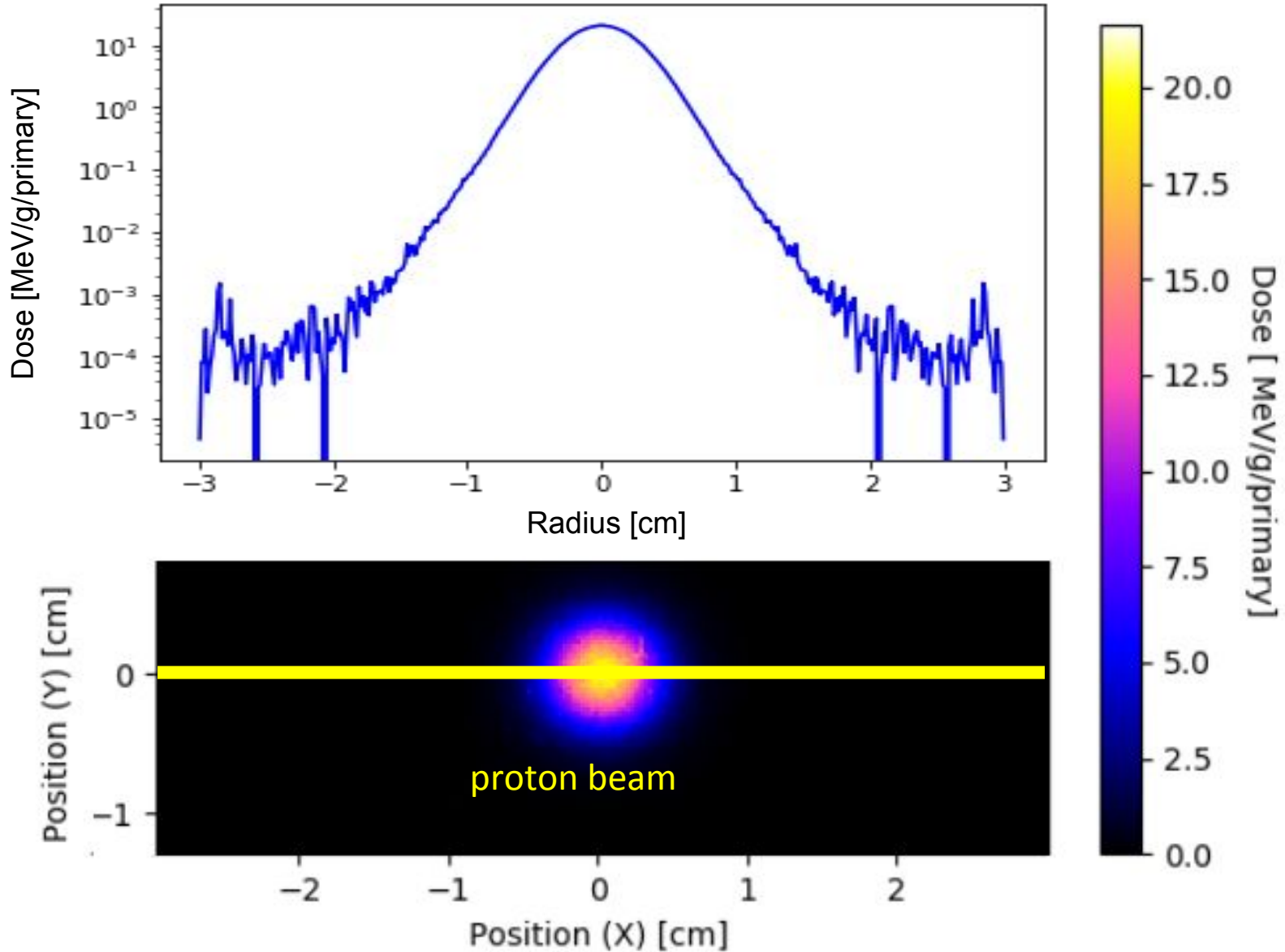
http://shieldhit.org/

authors:
    Niels Bassler - AU - Project lead
    Leszek Grzanka - IFJ-PAN / AU
    Nikolai Sobolevsky - INR RAS

- pymchelper - pymchelper.readthedocs.io

- generatemc - mcpartools.readthedocs.io
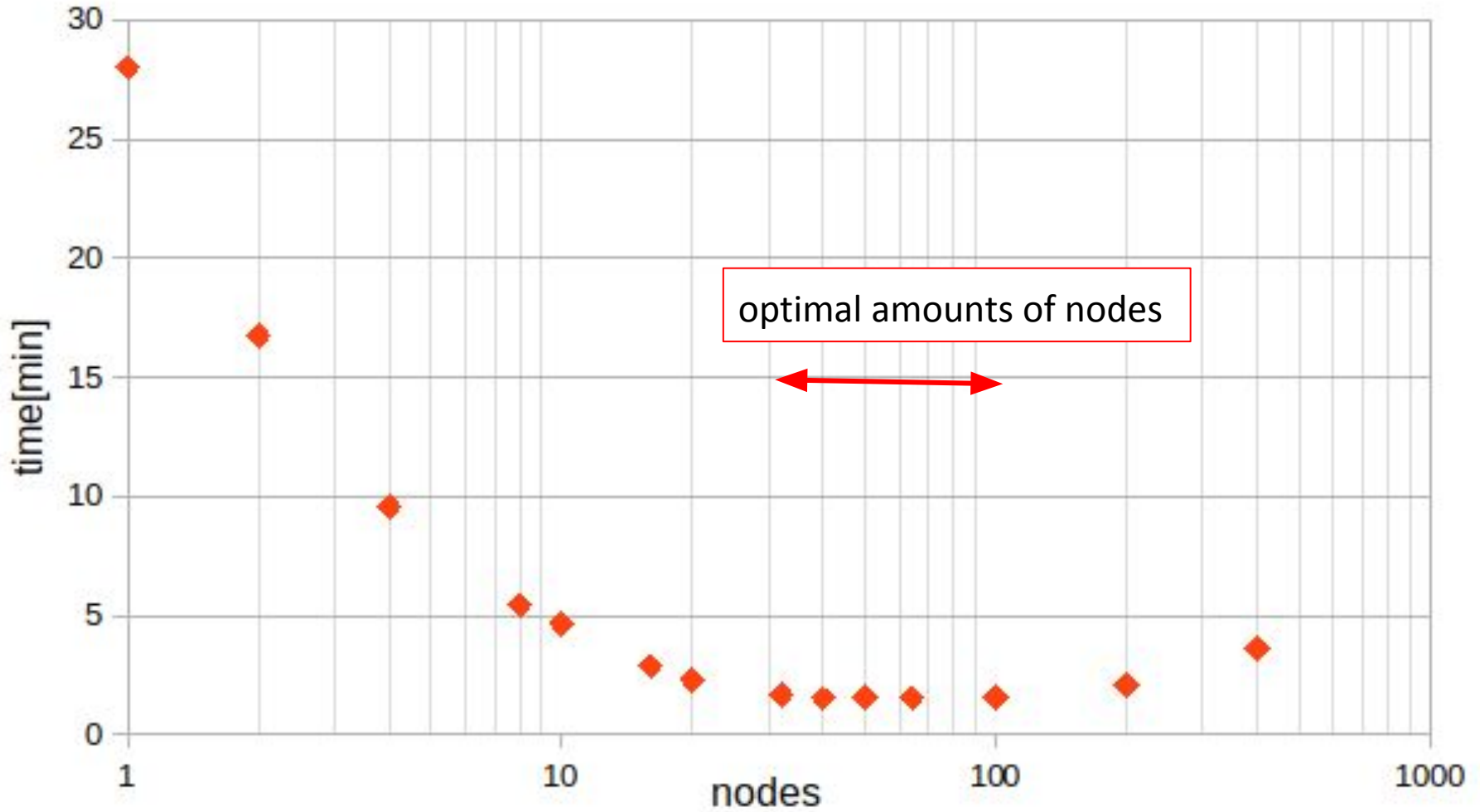
- convertmc - mcpartools.readthedocs.io

Simulated pencil beam dose distribution

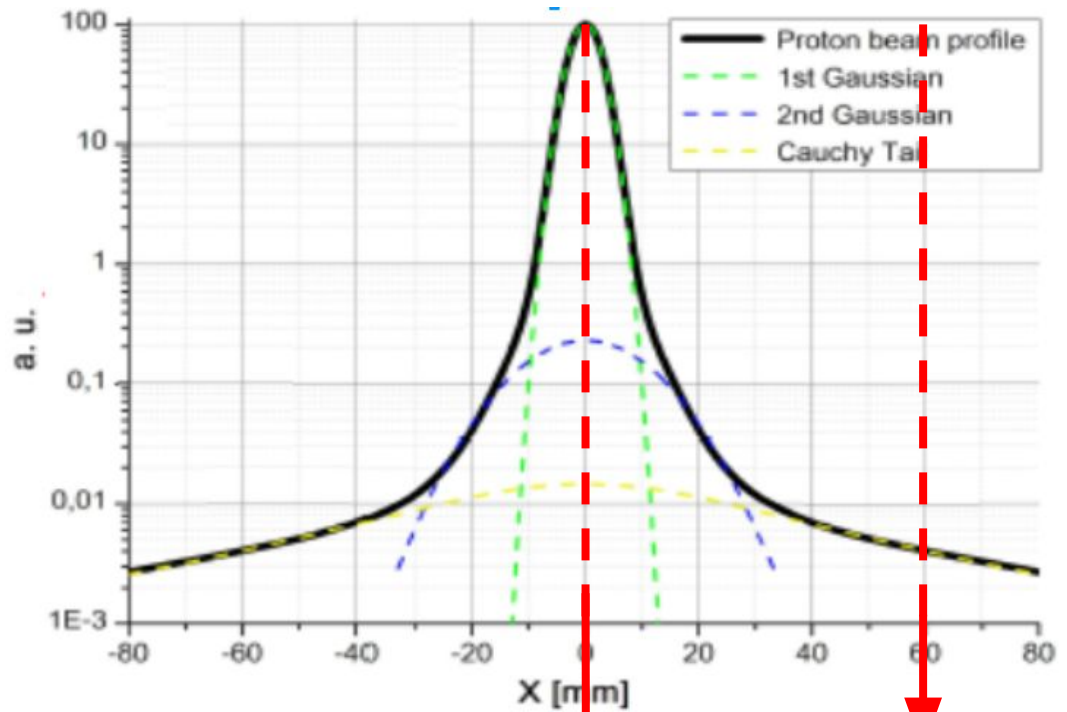# Optimization of MC calculation time on multi-node batch system
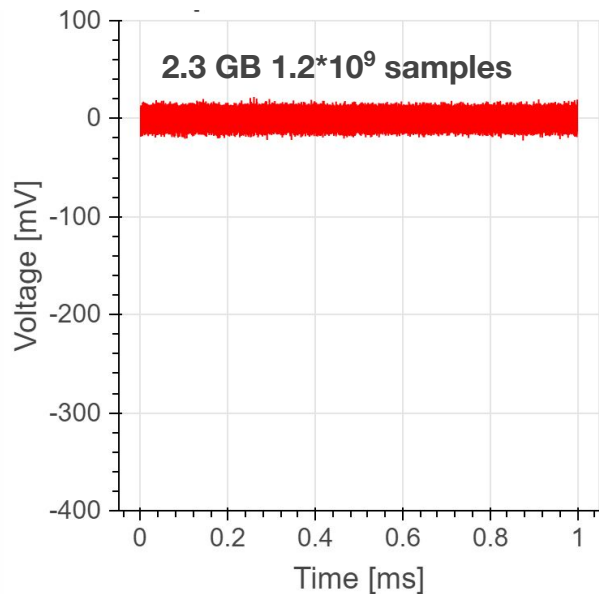

1 profile on private CPU = 15 hours
1 profile on PROMETHEUS = only 1 hour of simulation!
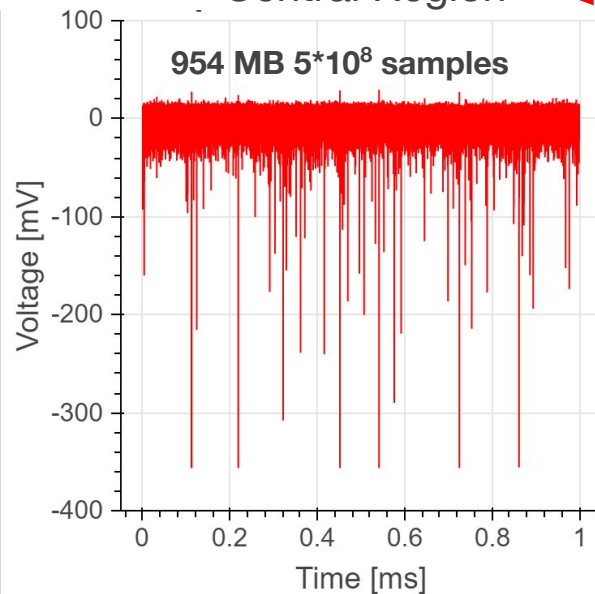
Data acquisition
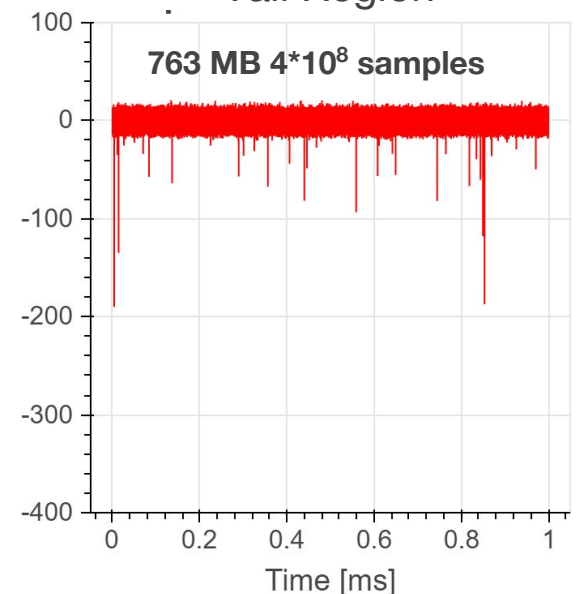
Elimination of noise from diamond detector signal

Statistical analysis enables data to be reduced by a factor of 10^4.

91 GB
$4.5*10^{10}$ samples

Threshold level

SIGNAL

NOISE

Counts

Voltage [mV]
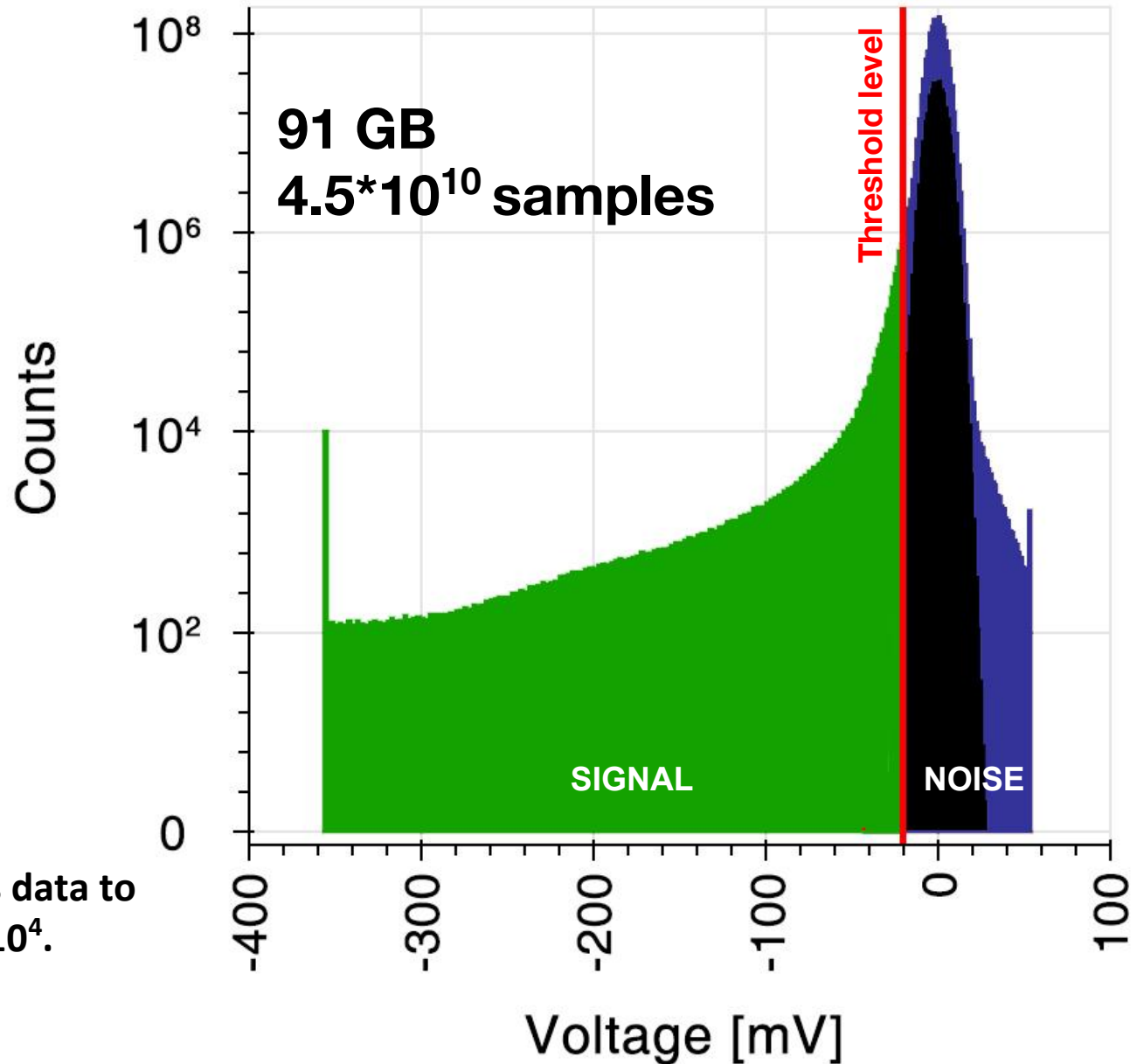
Single node, interactive mode, reading optimization

Reading into memory ~100 GB from lustre filesystem.

Speedup saturates at 7 cores

Node I/O as main bottleneck.

Sequential	VS	Parallel
52.5 min		7.5 min

Resources optimisation for parallel data processing
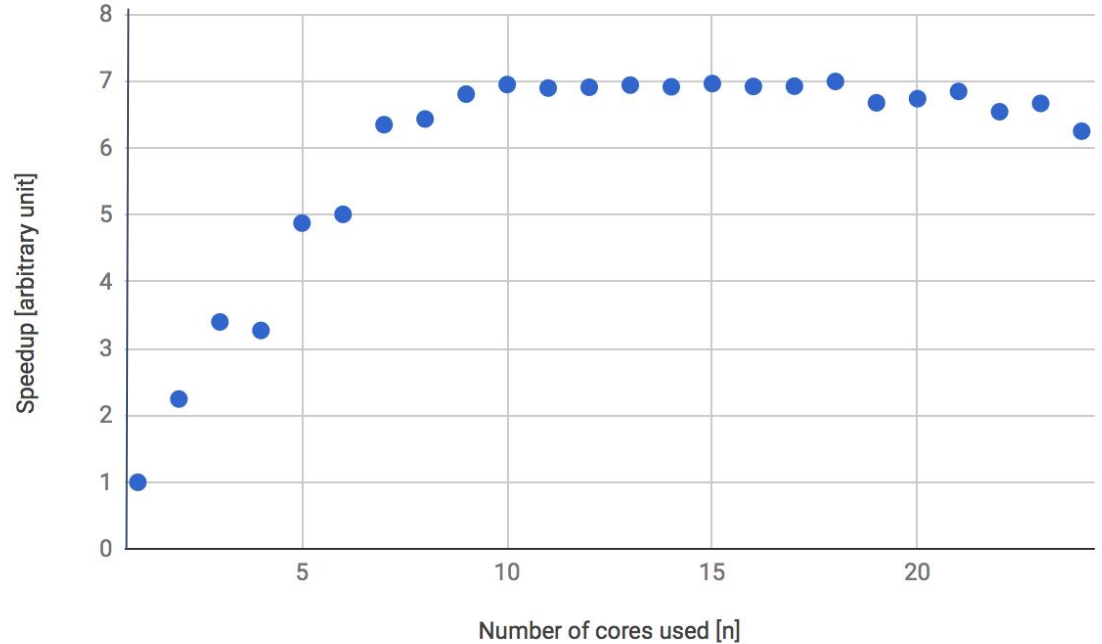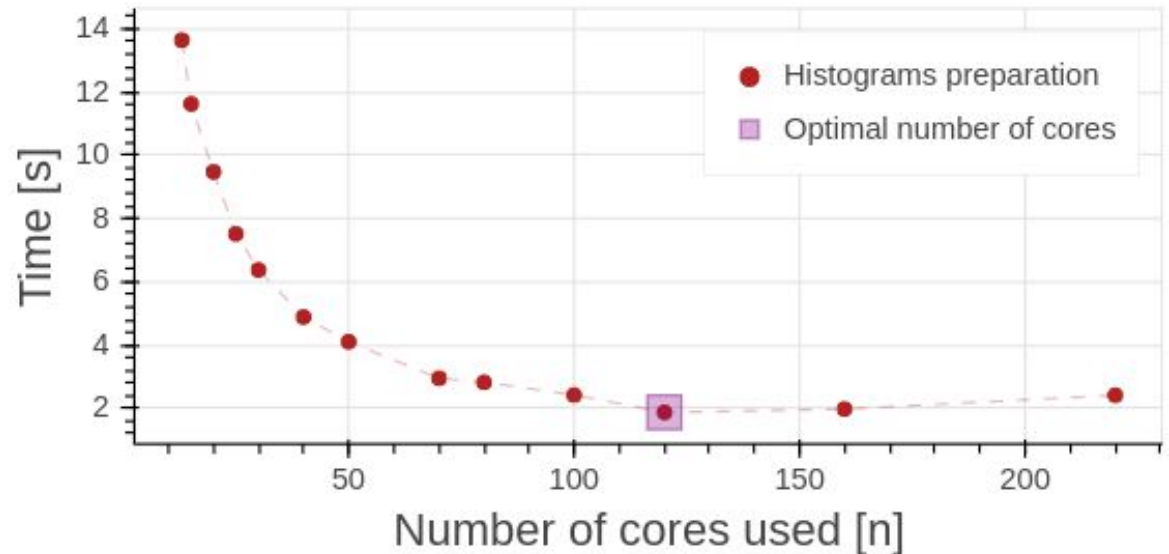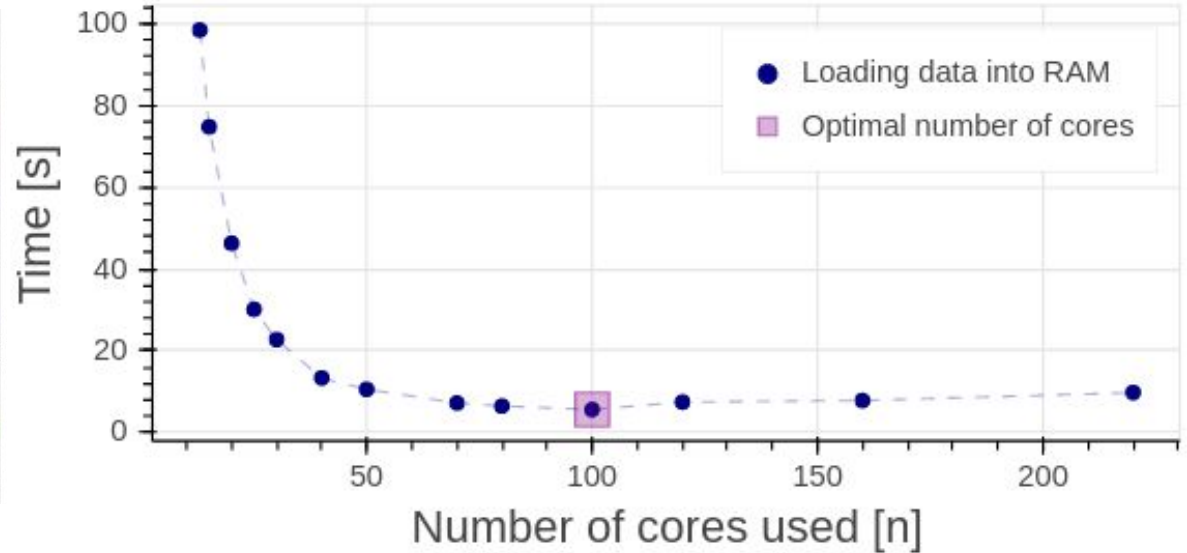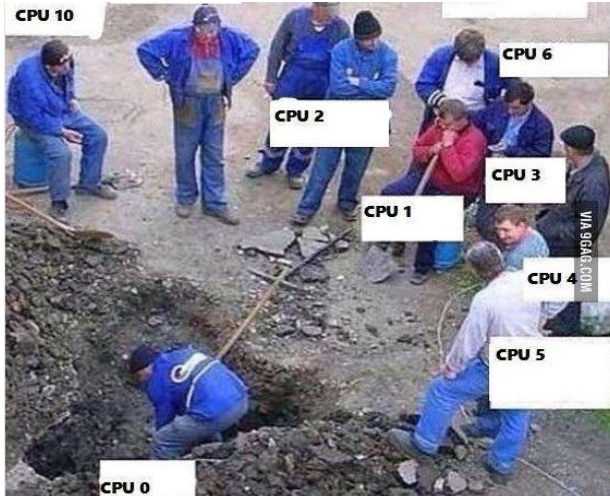

When processing 45GB of data, acceptable performance for on-line analysis require at least 100 cores for loading data into RAM and 120 cores to prepare histograms.

THE SOFTWARE AND THE STORAGE /COMPUTE RESOURCES

Software stack:
Slurm, Jupyter,
Interactive Batch,
Pro-Viz, SSH Tunneling,Python
Numpy, **Dask,** Pandas,
Apache Parquet, and Bokeh.

The total time needed to perform the simulation was around **10^4 h.** Storage resources used during **data analysis were 100-1000 GB**.
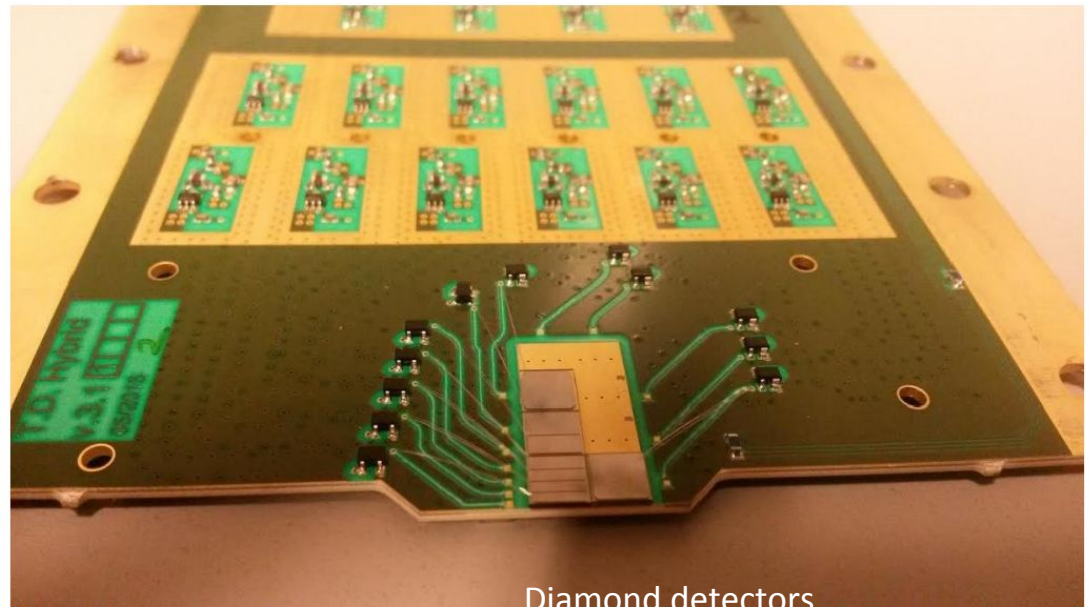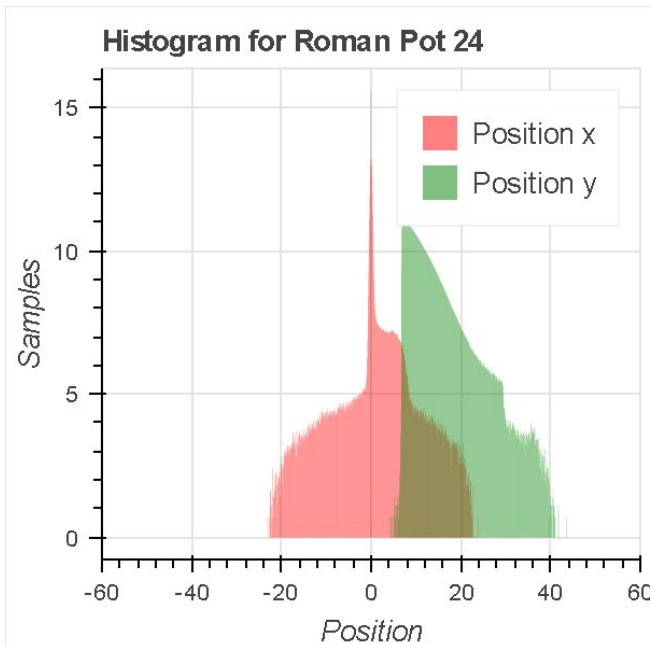
TOTEM at CERN

The **'Total, elastic and diffractive cross-section measurement'** experiment studies forward-emitted particles from LHC beam collisions.

Histogram for Roman Pot 24

Diamond detectors

~50 GB data volume

Summary

Simulation of diamond detector response:
- ❏ Particle transport code SHIELD-HIT12A used
- ❏ Parallel simulation using slurm batch system
- ❏ Optimum number of nodes selected to minimize total CPU time.

Proton beam data analysis (IFJ PAN + TOTEM):
- ❏ Estimation of threshold value to separate signal from noise
- ❏ Adoption of software stack for interactive data analysis, based on dask library
- ❏ Allocation of optimum resources to minimize data loading and processing time
- ❏ Proof-of-concept application for analysis of data generated by the TOTEM experiment

Conclusions

- The infrastructure of Prometheus and its available software stack have been succesfully used for interactive data analysis and Monte-Carlo simulation of the diamond detector data.
- Interactive analysis of large volume of data requires careful optimization of computing resources.
- Beam data analyses in radiation therapy and High Energy Physics pose similar computational challenges.

Thank You for attention!

This research was supported in part by PL-Grid Infrastructure.

Authors acknowledge grant no. MNiSW DIR/WK/2017/07