

Marcel Trattner

# Pulsar Data Analysis

A use case to increase scalability



© SKAO

# Big Data Challenges

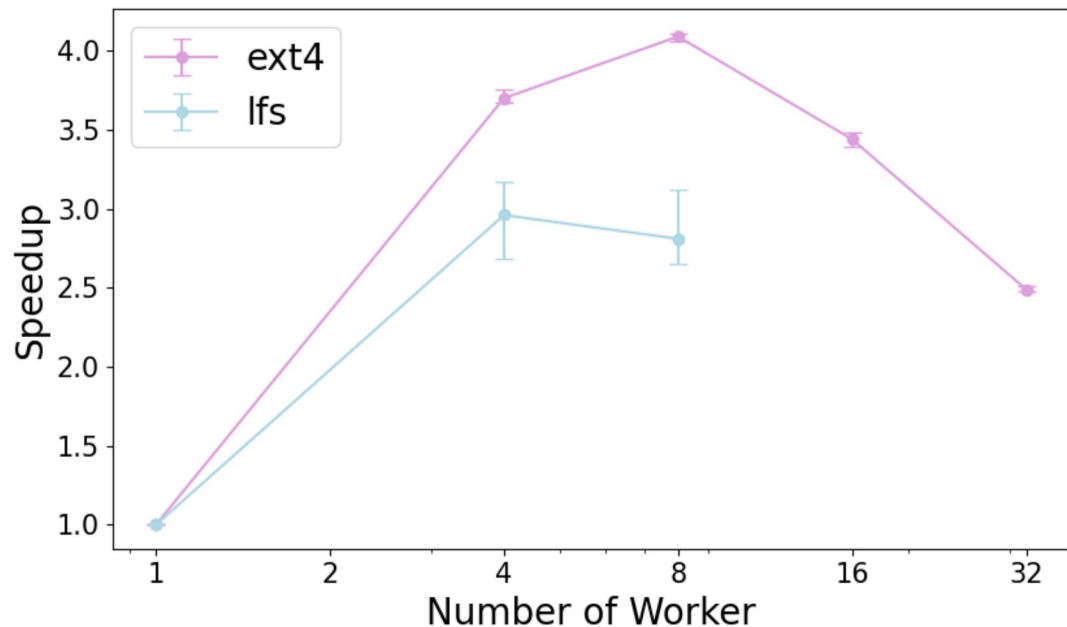
- Square Kilometre Array Observatory (SKAO)
- ~1 Petabyte per single image
- Novel demands on computing



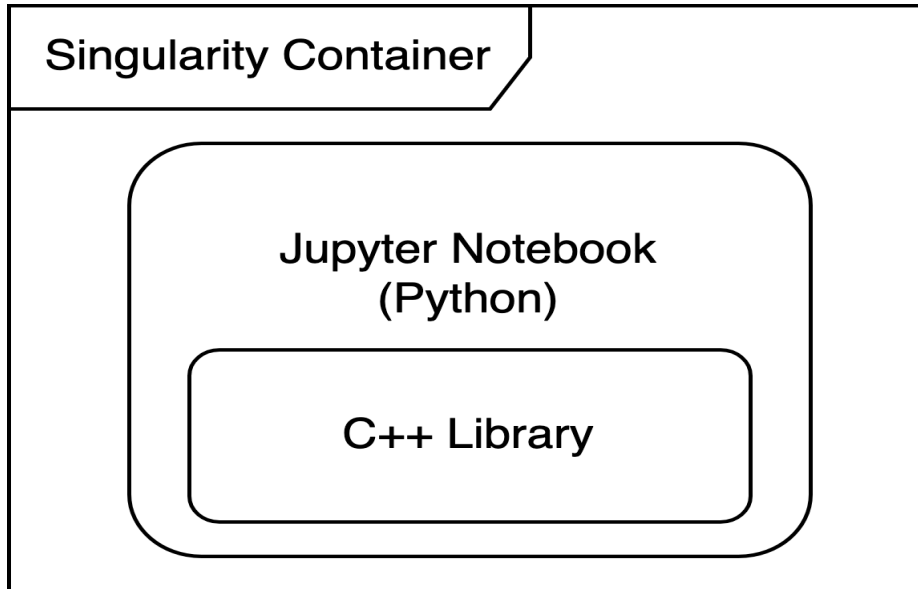
[https://skach.org/wp-content/uploads/2022/01/White-paper\\_Swiss-interest-and-contribution-in-SKA-1.pdf](https://skach.org/wp-content/uploads/2022/01/White-paper_Swiss-interest-and-contribution-in-SKA-1.pdf)

# Limitations in Scalability

- Parallel processing analysis of the astronomical framework CASA
- Relative speedup measured in relation to number of workers
- Maximal speedup has been reached at 8 nodes and declines afterwards



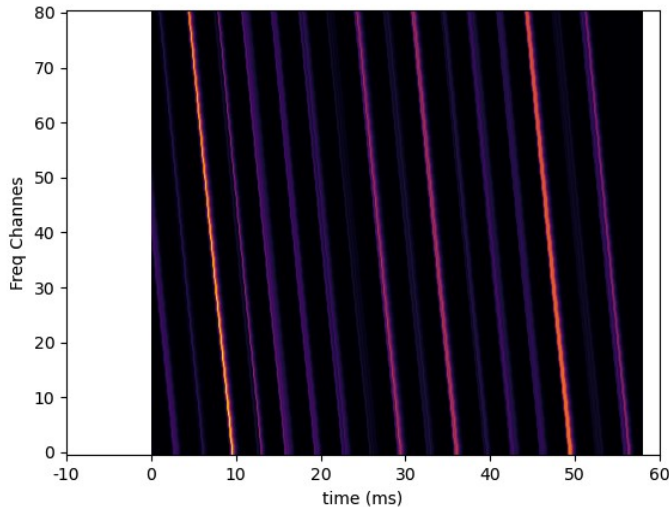
# Architecture of Pulsar Analysis Tool



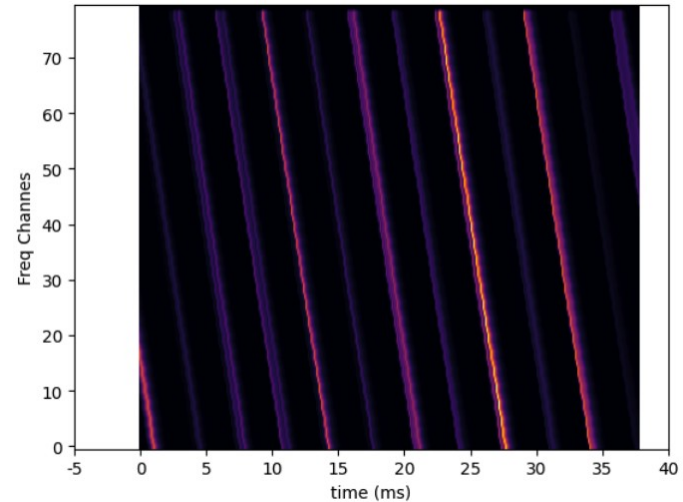
- Main code written in C++ for maximal performance
- C++ library embedded into python for easier usability through Jupyter Notebook
- All necessary dependencies packages inside Singularity container for fast deployment in data centers

# Implementation of Pulsar Simulation

Simulation of Pulsar data is at first conceptualized in Python and then translated to C++ for performance

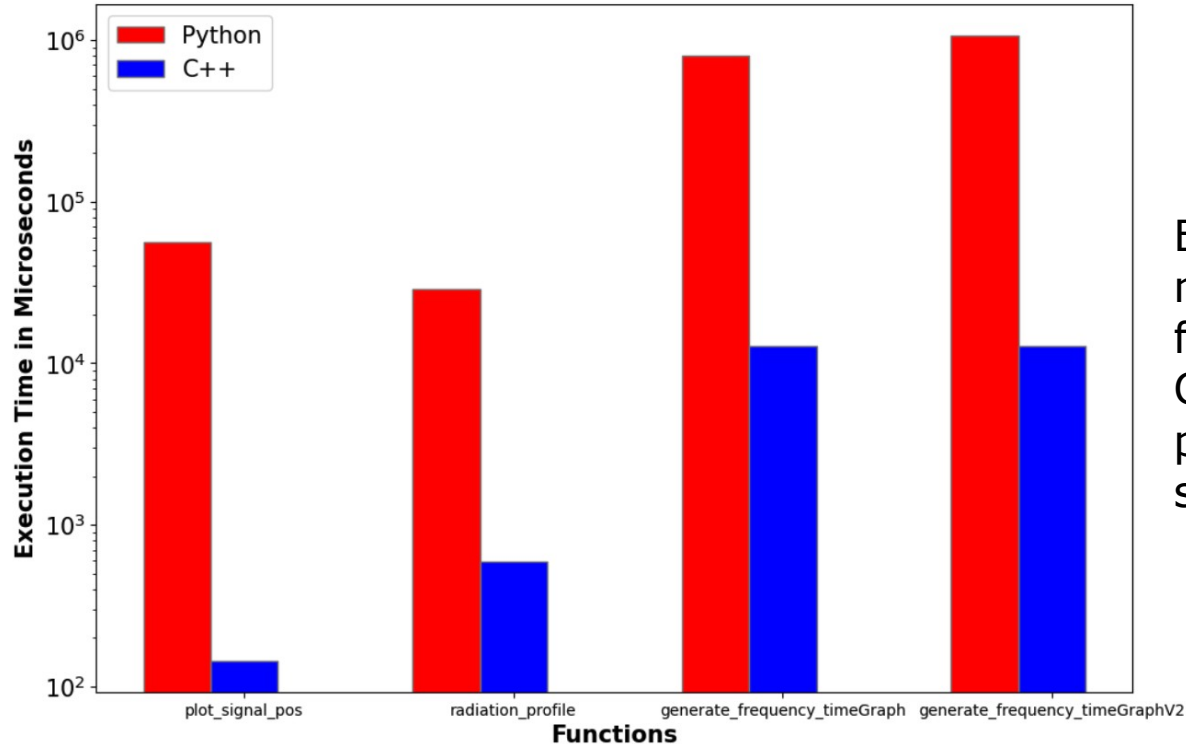


Python simulation:  
[Saha, 2023]



C++ translation:  
[Trattner, 2023]

# Python vs C++ Execution Time



Execution time in microseconds for different functions in the Python and C++ implementation are plotted on a **logarithmic** scale

# Next steps:

## Making Decisions for Architecture

- How can performance in parallel processing scale up?
- How can we distribute data chunks to individual workers?

Possible ideas:

- Shared Memory
- Memory Centric (50 TB addressable by HP)
- Lock Free?
- Steaming?

...



**Hochschule für Technik  
und Wirtschaft Berlin**

**University of Applied Sciences**

[www.htw-berlin.de](http://www.htw-berlin.de)