Contribution ID: **93**   Type: **Short Talk**

Detecting pulsar signals in vast real-time data streams with a machine learning / digital twin-based pipeline

Tuesday, 1 October 2024 17:10 (15 minutes)

One of the main benefits of modern radio astronomy, its ability to collect more higher-resolution and wider-bandwidth data from more and more antennas is now also starting to become one of its greatest problems. The advent of cutting-edge radio telescopes, such as MeerKAT, a precursor to the Square Kilometre Array (SKA), has made it impractical to rely on the traditional method of storing the raw data for extended periods and then manually processing it. Furthermore, the high data rates necessitate the use of High-Performance Computing (HPC), yet existing common radio astronomical data reduction tools, like Common Astronomy Software Applications (CASA), are not well-suited for parallel computing. We have addressed these challenges in developing the ML-PPA (Machine Learning-based Pipeline for Pulsar Analysis). It is an automated classification system capable of categorizing pulsar observation data and assigning labels, such as "pulse", "pure noise", or various types of Radio Frequency Interference (RFI), to each time fragment, represented as a 2D time-frequency image or "frame". The analysis is performed by a Convolutional Neural Network (CNN). Given the highly imbalanced distribution of different frame types in real data (e.g. only 0.2% are "pulses"), it is essential to generate artificial data sequences with specific characteristics to effectively train such systems. To achieve this, "digital twins" were developed to replicate the signal path from the source to a pulsar-observing telescope. A corresponding pipeline was created and tested in Python, and then rewritten in C++, making it more suitable for HPC applications. The initial version of the ML-PPA framework has been released and successfully tested. This talk presents a comprehensive overview of the project, its current status and future prospects.

Topic

Needs and solutions in scientific computing: Digital Twins

Primary authors: Mr KAZANTSEV, Andrei (MPIfR Bonn); Prof. BERTOLDI, Frank (University of Bonn); Dr DANGE, Gautam (FIAS Frankfurt); Mr TRATTNER, Marcel (HTW Berlin); Dr SAHA, Tanumoy (HTW Berlin); Mr OELKERS, Tim (HTW Berlin); PIDOPRYHORA, Yurii (MPG - Max-Planck-Gesellschaft); Prof. HESSLING, Hermann (HTW Berlin)

Presenter: PIDOPRYHORA, Yurii (MPG - Max-Planck-Gesellschaft)

Session Classification: Replicating and predicting complex systems with scientific Digital Twins