Detecting galaxies in terra-byte-sized HI spectral cubes

Narendra Nath Patra(1,2), Abinash Kumar Shaw(3,4), Manoj Jagannath(5), Aishrila Mazumder(1), Arnab Chakraborty(6), Rajesh Mondal(7), Samir Choudhuri(8)

(1) Department of Astronomy, Astrophysics & Space Engineering, Indian Institute of Technology Indore, Indore 453552, India, e-mail: naren@iiti.ac.in
(2) Astronomy and Astrophysics Division, Raman Research Institute, Sadashivanagar, Bengaluru - 560080, India.
(3) Department of Physics, Indian Institute of Technology Kharagpur, Kharagpur 721302, India.
(4) Astrophysics Research Centre, Open University of Israel, Ra’anana 4353701, Israel.
(5) Department of Electrical and Electronics Engineering, PES University, Bangalore 560085, India.
(6) Department of Physics and McGill Space Institute, McGill University, Montreal, QC, Canada H3A 2T8.
(7) Department of Astronomy and Oskar Klein Centre, AlbaNova, Stockholm University, Stockholm SE-10691, Sweden.
(8) Department of Physics, Indian Institute of Technology Madras, Chennai, 600036, India.

The upcoming Square Kilometer Array (SKA) is expected to produce a humongous amount of data for undertaking HI science. We have developed an MPI-based Python pipeline to deal with the large data efficiently with the present computational resources. Our pipeline divides such large HI 21-cm spectral cubes into several small cubelets and then processes them in parallel using the publicly available HI source finder SoFiA-2. The pipeline also takes care of sources at the boundaries of the cubelets and also filters out false and redundant detections. By comparing with the true source catalog, we find that the detection efficiency depends on the SoFiA-2 parameters, such as the smoothing kernel size, linking length, and threshold values. We find the optimal kernel size for all flux bins to be between 3 to 5 pixels and 7 to 15 pixels, respectively, in the spatial and frequency directions. Comparing the recovered source parameters with the original values, we find that the output of SoFiA-2 is highly dependent on kernel sizes and a single choice of the kernel is not sufficient for all types of HI galaxies. We also propose the use of alternative methods to SoFiA-2, which can be used in our pipeline to find sources more robustly.