4th URSI AT-RASC, Gran Canaria, 19 – 24 May 2024

Imaging Spectral-Line Deep Fields in the SKA-Era: insights from CHILES and DINGO

R. Dodson*(1), M. Meyer(1), D.J. Pisano(2), A. Williamson(1), P.J. Elahi(3), J. Rhee(1), K. Vinsen(1), Q. Gong (4) and K. Rozgonyi(1)

(1) ICRAR, M468, The University of Western Australia, 35 Stirling Hwy, Crawley, Australia 6009

(2) University of Cape Town, Rondebosch 7700, South Africa

(3) Pawsey Supercomputing Research Centre, CSIRO, Perth, Australia

(4) Oak Ridge National Laboratory, USA

1 Introduction

To measure the HI mass of a large range of galaxy masses over a large range of redshifts it is essential to have deep HI surveys, where many epochs are combined to provide an ultra-sensitive observations. Deep Spectral Line imaging with SKA and the pathfinders is one of the greatest challenges for the next generation instruments as the data volumes are too large to store the raw visibilities to be used for forming the final image. The SKA current plans are for the instruments to have sufficient storage for one day of spectral line data; deep imaging will then be formed from many days of observations. This is predicated on the assumption that the images have Gaussian residual errors and no systematic errors – which is rarely the case.

Deep Investigation of Neutral Gas Origins (DINGO) has been awarded 3,200 hours over about ∼400 days, on the Australian SKA Pathfinder (ASKAP). Initially DINGO was following the SKA strategy. COSMOS HI Large Extragalactic Survey (CHILES) uses the J-VLA, and has 1000 hours over 5 epochs in B-array[2], made up of several hundred individual observing sessions. The total data was fully retained. The CHILES analysis shows that it is vital to retain the visibilities, at least in some form, to address unexpected systematic effects. In our case the sensitivity was so high that sources out to 3^o from the phase centre had significant impact of the residual image.

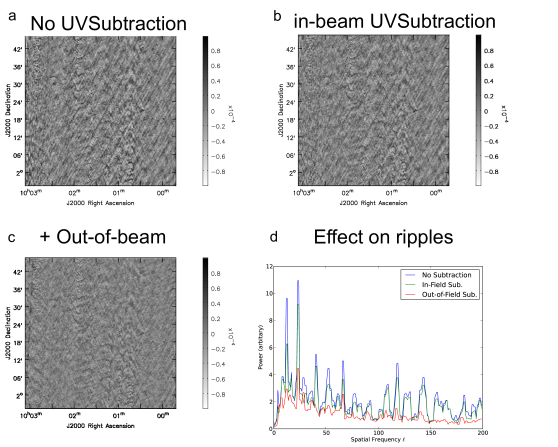

Figure 1. Effect of correcting for residuals in the image cube; a) without extra uvsubtraction, b) subtraction of in-beam sources, c) with subtraction of out-of-beam sources and the d) effect in the Fourier domain [2].

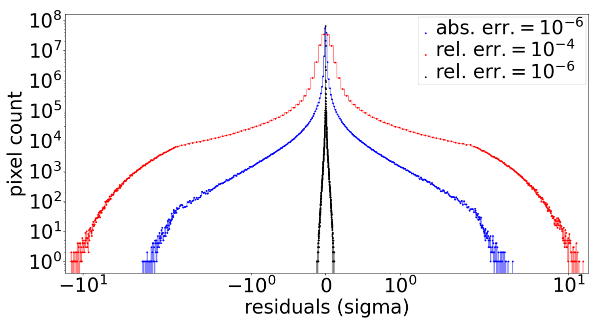

Figure 2. Histograms of residual errors introduced by the lossy MGARD compression the image cubes, relative to no compression. Three methods are shown: absolute errors less than 10^{-6} (blue) and relative errors less than 10^{-4} (red) and 10^{-6} (black). These have worst case errors of about 5, 10 and 0.1σ and compression ratios of 18, 24 and 13, respectively.

This is informing our work on DINGO[3] where we cannot store the full 30PB datasets. Now we store the daily datasets as gridded data[4]. These are (potentially) a similar size to the images, have the full advantages of optimum Baseline Dependent Averaging – whilst correctly applying the kernels, allow compression (both lossless and lossy) using ADIOS2/MGARD embedded in the CasaCore libraries[5] and are showing excellent performance.

We are now entering full survey operations, already with more than 100 hours or 15 nights of data (in the combined pilot and survey data) on target field G23 ($22^h59^m00^s$, $-32^o00^m00^s$). We are combining these using compressed grids, that show a seven-fold improvement in processing time, a six-fold (lossless) or better (∼20 for a lossy error bound of 5σ) reduction in storage requirements, with minimal impact on the results.

References [1] Hopkins & Beacom, 2006, ApJ, 651, 142 [2] Dodson, R. etal., 2022, AJ, 163, 59 [3] Rhee, J. etal., 2023, MNRAS, 518, 4646 [4] Rozgonyi, K., 2020, PhD, UWA [5] Gong, Q., etal., 2023, Software-X, 24, 101590