URSI GASS 2021, Rome, Italy, 28 August - 4 September 2021

Testing a New Radio Interferometry Data Reduction Framework

Jan-Willem Steeb* [(1)], Ryan Raba[(1)], and Andrew McNichols[(1)]
(1) National Radio Astronomy Observatory, Charlottesville VA, USA

An ambitious program has been undertaken to develop scalable, maintainable, Python-based software suitable for the next generation of radio telescopes using off-the-shelf data structures and processing frameworks [1]. Prototype functions are composed using Dask, Numba, Xarray, and Zarr [2, 3, 4, 5] to be functionally equivalent re-implementations of a subset of tools from the Common Astronomy Software Applications (CASA) [6].

The CASA mosaic gridder [7] used by the ALMA project was taken as representative of current imaging capability. For a larger than memory ALMA dataset on a single-machine the prototype's performance improves proportional to the number of data partitions until communication overhead begins to dominate the run time. The computation on the smaller partitions can be more efficiently organized by the Dask scheduler. Thus far, tests of the imaging prototype indicate comparable or improved performance with respect to existing implementations, and greater potential for rapid horizontal scaling.

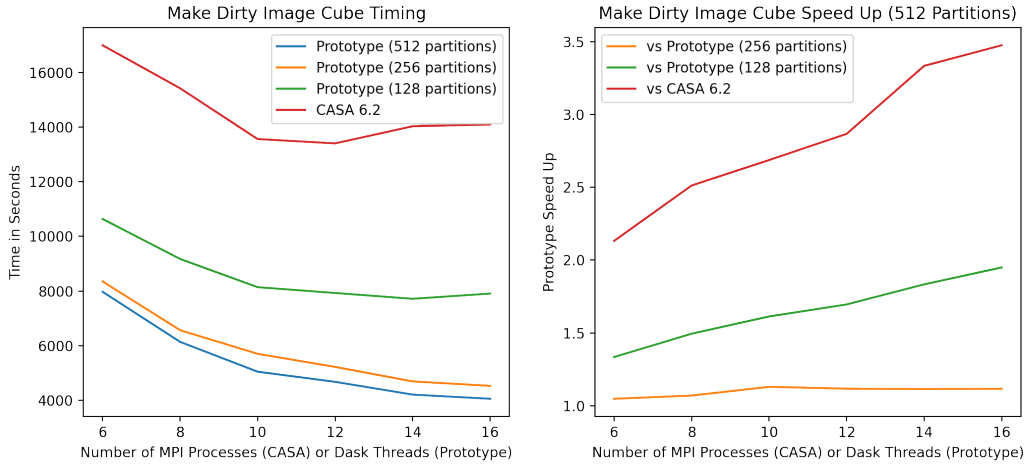

Figure 1. Run time and speedup as a function of parallelism. An ALMA dataset was used with 27 pointings and 16284 channels (367 GB of visibility data).

References

[1] CASA Development Team, "CASA Next Generation Infrastructure," 2021, `https://cngi-prototype.readthedocs.io`.

[2] Dask Development Team, "Dask: Library for dynamic task scheduling," 2016, `https://dask.org`.

[3] S. Lam, A. Pitrou, and S. Seibert, "Numba: a LLVM-based Python JIT compiler," *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, November 2015, pp. 1-6, doi:10.1145/2833157.2833162.

[4] S. Hoyer, C. Fitzgerald, J. Hamman, et al., "Xarray: v0.8.0," 2016, doi:10.5281/zenodo.59499

[5] Zarr Development Team, "Zarr: An implementation of chunked, compressed, N-dimensional arrays for Python," 2016, `https://github.com/zarr-developers/zarr-python`.

[6] J. P. McMullin, D. Schiebel, W. Young, and K. Golap, "CASA Architecture and Applications," *Astronomical Data Analysis Software and Systems XVI (ASP Conf. Ser. 376)*, 2007 , pp. 127–130.

[7] CASA Development Team, "Synthesis Imaging," `https://casadocs.readthedocs.io/en/latest/notebooks/synthesis_imaging.html`.