

# Data Rates for the ALMA Archive and Control System

*Steve Scott, Steve Myers*

**7 Oct, 2000**

- 1) The ALMA data rates assume a 64 antenna array. All data rates scale with the number of baselines and will thus scale with the number of operational baselines as the array is constructed.
- 2) The average data rate is 6 MB/sec. This data rate should be technically feasible without great difficulty by the time that ALMA is completed. This average data rate should be the average over long periods of time and can be used to determine archive size. Some projects will use data rates higher than the average and some lower.
- 3) The peak data rate will be 60 MB/sec. It may be necessary to stage high rate data to intermediate storage before archiving.
- 4) The archived data shall consist of the visibility data, images, monitor data, and the scripts used to collect and reduce the data. The visibility data and images will comprise the majority of the data (>95%). When long integrations are used for the images, the images will be stored in the archive. When shorter integrations are used, the images will be generated on the fly from the visibilities upon extraction from the archive. The break point between these two techniques will be determined by the computing capability of the archive extraction pipeline, and may evolve over time. To ensure that images are always available from the ALMA archive for all projects, images must always be archived if the pipeline cannot generate them upon extraction.
- 5) Visibility data that has been corrected for atmospheric phase effects shall be available as well as the uncorrected data. In the early years of ALMA, both will be archived. When ALMA is a mature instrument, it will ideally automatically choose the best, on an antenna or perhaps baseline basis. The user shall in principle be able to select whether to archive corrected data, uncorrected, both, or an automatic choice of the best. This selection shall apply to all baseline data.
- 6) The corrected and uncorrected visibility data will be integrated over the same time periods. All baselines will be integrated over the same time periods.
- 7) The user shall be able to specify the number of spectral channels and integration time to meet the science goals. These choices combined with the corrected/uncorrected selection will yield a data rate. For example, a project with 1850 spectral channels, 10 second integration time, archiving both corrected and uncorrected data will meet the average data rate of 6MB/sec. This assumes a complex visibility is stored as two 16-bit scaled integers, an image pixel is stored as a 32-bit integer, and images are produced about every 5 minutes (data volume evenly split between images and visibilities).
- 8) The proposal preparation tool shall calculate data rates and total data volume for a project.
- 9) The science, scheduling, or operations group shall determine the policies and methods (if any) of allocating and enforcing data rates for projects. They may also have restrictions on the allowed combinations of corrected and uncorrected data (for example, they may always require recording of both), and these restrictions may change over time.