ALMA SSR Requirements for the AIPS++ Benchmark Effort

Version: May 28, 2004

The ALMA Science Software Requirements (SSR) committee has identified four requirements relating to benchmarking AIPS++ performance:

- **2.1.1 R4 -** The performance of the Package shall be quantifiable and commensurate with the data processing requirements of ALMA output and the scientific needs of users at a given time. The timing and reproducibility of results for a fiducial set of reduction tasks on specified test data will be benchmarked and compared against other packages and a list of benchmark specifications shall be provided and maintained by the Project.
- **2.2.2 R1.1 -** The GUI shall provide real-time feedback via standard compact displays. Window updating must be fast (less than 0.1s on same host).
- **2.3.2 R4 -** The Package must be able to handle, efficiently and gracefully, datasets larger than main memory of the host system.
- **2.7.2 R3 -** The [display] plot update speed shall not be a bottleneck. Speed shall be benchmarked, and should be commensurate with comparable plotting packages.

The development of a benchmark program requires close collaboration between the SSR and AIPS++ developers. In particular, the SSR must provide guidelines for the benchmark process, data sets to benchmark, comparison scripts in other packages, and oversight to ensure that processing speed is adequate to meet ALMA requirements. The AIPS++ developers must provide dedicated benchmark machines, run the scripts in each package and publish the results on the web. In addition, once benchmark numbers are quantified and areas are identified that need improvement, SSR and AIPS++ developers set priorities for future work that will minimize processing speed in specific areas. This document specifies development guidelines and expectations for the development of a benchmark program for AIPS++.

Phase I: ALMA Benchmark infrastructure development and first comparisons

This phase of the benchmark process is designed to satisfy part of the requirement **2.1.1 R4**. At the end of this phase, the performance of the AIPS++ package shall be quantifiable for a limited number of datasets. The timing and reproducibility of results for a fiducial set of reduction tasks on specified test data will be benchmarked and compared against other packages and a list of benchmark specifications shall be provided and maintained. It is also designed to test whether the AIPS++ package can handle, efficiently and gracefully, datasets larger than the main memory of the host system (e.g. requirement **2.3.2 R4**).

AIPS++ performance shall be compared to GILDAS/CLIC (used for IRAM & Plateau de Bure data), MIRIAD (used for BIMA & ATA data), and AIPS (used for VLA data). A single comparison between all packages is not possible due to restrictions in the type of data format that can be filled into each package:

There is no current task in GILDAS which can convert PdBI data format to standard FITS. Thus, we can compare processing speed for interferometric data between AIPS++ and GILDAS

Package	Standard FITS	ALMA-TI FITS	AIPS++ format	PdBI format	MIRIAD format	AIPS archive format
GILDAS		√		√		
MIRIAD	√				√	
AIPS	√					√
AIPS++	√	√	√			√

(filling ALMA-TI FITS format data generated by GILDAS), and between AIPS++, MIRIAD, & AIPS (filling FITS format data).

Note: while GILDAS and AIPS++ can reduce single dish data, MIRIAD and AIPS cannot. Note also, GILDAS is restricted to processing single dish data in IRAM format only, and there is no conversion algorithm that will convert this format to, e.g., FITS or ALMA-TI FITS. Because of this format discrepancy, benchmark comparisons for single dish data processing cannot be done at this time. This plan will be updated to include benchmark tests for single-dish processing at a later time.

Benchmark Dataset Descriptions

The SSR selected two datasets initially to obtain first comparisons and develop the benchmark infrastructure with automatic web publishing. Data and initial scripts were provided by the SSR. Detailed comparison of the script steps were made by AIPS++ to ensure that we are comparing "apples-to-apples." Each dataset has specific goals which quantify different aspects of system performance:

- **Pseudo GG Tau data:** PdBI data of 25 March that has been expanded to 64 antennas. The source structure has been converted to a point source. The data include simultaneous 3 & 1 mm continuum and spectral line emission. Data is provided in ALMA-TI FITS format; reduction/imaging can be compared with GILDAS only. This is the same data that was used during the AIPS++ "Phase III" test to get a first complete snapshot of the speed of AIPS++ processing relative to GILDAS. Goals:
  - Ensure continuous comparisons between AIPS++ performance improvements with time and the "Phase III" results.
  - Obtain realistic comparisons of run-time processing of core functions (e.g. filling, calibration, imaging) on an ALMA-size dataset. Processing time required for each step will be dominated by actual data access or processing functions, not by initial setup.
  - Exercise millimeter-specific processing steps: e.g. polynomial fit of the gain solutions with time; phase solution transfer from 3 mm to 1 mm during gain calibration.

  Issues & Limitations:
  - Cannot compare imaging of extended structure (e.g. requiring clean-deconvolution regions).
  - The Export Data Format for ALMA has not been defined, thus, the performance comparison of the ALMA-TI FITS filler, while interesting, may not be directly relevant.

ALMA-TI FITS filler performance in AIPS++ will not be optimized since this format will be obsolete soon. Filler priorities will be established once the ALMA export data format has been defined and a filler function is implemented.

  - No polarization processing possible.
  - Self-calibration steps not exercised due to limited S/N.
  - Some processing steps do not have a 1-1 correspondence (e.g. *initialization*) and a comparison is not relevant for these steps.

- **Polarized continuum snapshot:** VLA polarized continuum emission in the gravitational lens 0957+561 at 6 cm wavelength in one spectral window. The dataset has been extended with the AIPS++ simulator to increase total integration time and, hence, processing time (so the comparison will not be dominated by setup tasks). The number of antennas (27) is not increased to excercise imaging tasks with extended emission. Data reduction/imaging will be compared with MIRIAD & AIPS. Goals:
  - Exercise full polarization calibration, self-calibration, non-point source imaging.
  - Obtain comparisons of run-time processing of core functions (e.g. filling, calibration, imaging) on a medium sized dataset. Processing time required for each step will be dominated by data processing time.

  Issues & Limitations:
  - Polarization processing can only be compared with MIRIAD & AIPS.
  - MIRIAD can only self-calibrate based in the total intensity image while AIPS++ does a full I,Q,U self-calibration. This difference must be kept in mind when comparing processing speed.

At the end of Phase I of the ALMA benchmark program there will be a dedicated benchmark machine with working AIPS++, GILDAS, MIRIAD, and AIPS builds. There will be a web page with first results of the comparisons for the two datasets described above. GILDAS, MIRIAD, and AIPS builds and benchmark numbers will remain static unless there is an improvement in the code-base; at which time, new software versions will be installed on the benchmark machines and the scripts will be re-run. New AIPS++ numbers will be generated with each stable build to track performance improvements. Based on the performance comparison, the SSR subsystem scientist, ALMA management, and the AIPS++ project will establish the priorities for which areas of the code-base require further improvement.

Phase II: Data set expansion

Phase II of the ALMA benchmark program expands the number of datasets being compared to broaden the parameter space over which processing speed is compared and concentrate on comparing larger, more ALMA-like datasets.

Two datasets have been identified by the Offline subsystem and the SSR for further benchmarking:

- **Spectral line fast-switching data:** VLA NH_3(1,1) emission in the star forming region G192.16–3.84 at 1.3 cm (23.694 GHz) wavelength in one spectral window. There were 27 antennas and 127 channels in the spectral line cube. Editing is required to remove bad antennas at specific timestamps and to clip high points during some observations. The track was approximately 5 hours long, with initially unstable weather (e.g. poor phase stability) that later settled down. The fast-switching cycle time was 4 min. A 22 GHz tipping scan provided an estimate of the zenith opacity (0.062). Continuum subtraction in the *uv*-plane is necessary to remove an unresolved continuum source near the image phase center. Source structure is complex with both weak ($2 - 3\sigma$) and stronger (10σ) emission. Data reduction/imaging will be compared with AIPS (& MIRIAD?). Goals:
  - Exercise calibration on fast-switching spectral-line data with rapid phase variations, opacity correction, *uv*-continuum subtraction, non-point source imaging.
  - Obtain comparisons of run-time processing of core functions (e.g. fill, edit, calibration, split, imaging) on a large data cube (this dataset is significantly larger than the 0957+561 polarized continuum data and stresses read and write times of the measurement set and calibration tables). Processing time required for each step will be dominated by data processing and I/O time.

  Issues & Limitations:
  - When first filled, this dataset is 1.334 GBytes which then expands to 4.222 GBytes when MODEL, CORRECTED_DATA, and IMAGING_WEIGHTS columns are created. The final 256×256 image cube with 121 channels is $\sim$ 31 MB in size. Adequate memory must be available for processing.

- **Large mosaic data:** ACA HI mosaic of the Large Magalenic Cloud. Data reduction/imaging will be compared with AIPS & MIRIAD. Goals:
  - Exercise calibration a large mosaic data cube, mosaic, non-point source imaging.
  - Obtain comparisons of run-time processing of core functions (e.g. fill, edit, calibration, split, imaging) on a large mosaic data cube.

  Issues & Limitations:
  - TBD

The SSR will identify datasets in the following areas:

- Spectral line, polarized emission (VLA). Multi-configuration datasets if possible.
- Large, simulated dataset which include atmospheric opacity variations and phase noise.
- Single-dish + interferometer combination in the *uv* (and image?) plane (no single dish reduction, just image combination).

Phase III: Benchmarks affecting the User Interface

The AIPS++ Glish framework is being replaced with ACS/Corba. Glish-based GUIs will be replace with JAVA GUIs once the framework conversion is complete. Thus, benchmark comparisons which affect the GUI and plotting interface will be delayed until JAVA GUIs are ready to test. At this time, user tests identified in the ALMA Offline subsystem Test Plan shall quantify the GUI and plot performance (e.g. requirements **2.2.2 R1.1** & . **2.7.2 R3**).