ALMA Memo No. 606
ALMA & Allegro Phase Metrics Workshop Report

Authors : Luke T. Maud, Ed Fomalont, Remo Tilanus, Bill Dent

Summary

The ALMA long baseline signal propagation properties, mainly the variations of delays and phases among the antennas, were discussed in four-day meeting in Leiden, the Netherlands, under the auspices of the Allegro group. The presentations and discussions covered the various phase/delay instrumentation; the measurements of the delays above the array for various timescales between seconds to hours using carefully prepared interferometric test observations of quasars; the determination of astrometric results and antenna positions; and the lessons learned from the 2014 and 2015 Long Baseline Science Campaigns.

The major recommendations for improved future ALMA long-baseline are:

- The water vapour radiometers reduce substantially the short-term phase fluctuations. **Improvements in their performance and application algorithm should be continued**. The remaining phase fluctuations are caused by the dry air component which is extremely hard to model and monitor, even with many weather monitors around the array. New tests will be undertaken to investigate corrections for the dry term.

- Good quality imaging and little coherence loss requires short-term (i.e. the cycle time) RMS phase fluctuations no larger than 30° at the observing frequency—as monitored by short GonoGo or bandpass observations.

- **The short-term phase fluctuations (5-60 sec) do not increase much with baselines longer than about 2km**. However, the longer-term (> tens of minutes) and more systematic phase variations related to antenna positions do increase roughly linearly with baseline length;

- **Hence, the closer the phase calibrator is to the target, the better quality the image or derived astrometric position**. Cycle times between calibrator and target should be at most 2 min in the long baseline configurations. Only observations with sources closer than 2 to 3° can utilise the reduction in phase RMS when employing fast switching. Ideally, cycle times should be dynamic to account for source separation and weather conditions - unstable phase requires faster switching. **An updated GonoGo test cycling between a pair of quasars would provide a better test of atmospheric stability and also account for cycle times and source separations**.

- The future ALMA Long baseline (and high frequency) observations require improved monitoring of the data quality and phase stability during the observations or just upon completion in order to assess the viability of the observation as soon as possible.

Introduction

In late 2015 during the second Long Baseline Campaign (LBC), a decision was made to hold a workshop for the many people who worked heavily on the the ALMA long baseline observations between September 2014 to November 2014 and again between September 2015 to November 2015. During these periods, both PI observations and many tests were made. A main concern for long baselines observations is the phase stability, both the short- and longer-term systematic phase characteristics, which we designated as the *Phase Metrics* and *Phase Characteristics* of the LB data, respectively. The main concept of the workshop was to cover and discuss the many aspects of ALMA phase concerns in general, with emphasis on the tests and tentative results from the various groups involved, and to outline future directions and work needed.

Allegro, the Dutch ALMA arc-node, hosted the meeting in spring-summer 2016. It was held between 30 May and 3 June 2016 (for the list of over 20 attendees see Appendix A). The meeting was appropriately timed after the LBC campaign and before the start of Cycle 4 and therefore the forthcoming 2016/2017 LBC. Although the LBC in 2014 and 2015 proved to be 'successful', there are still problems to be addressed to improve the efficiency and data quality of future ALMA. This workshop focused on long-baseline phase metrics/characteristics and additional issues for high frequency observations at shorter baselines which have similar phase problems. The phase workshop promoted discussion and tighter working relationships with the leading researchers in the field looking towards the future tests and Cycle 4/5, but moreover, was the facility to discuss the outcomes and conclusions of the previous tests and to present recommendations based upon these experiences.

This document summarises the talks and many discussions during the workshop and indicates the recommendations that should be considered for the future long baseline science operations, engineering concerns and additional LBC/High Frequency testing. The broad discussion points were: (1) What have we learned from 5 and 10 km configurations, (2) What are the major limitations and *how* can they be addressed, and (3) understanding the effects of phase stability.

Meeting Topics and Talks

The workshop was arranged into topic areas, each with dedicated sessions during the week long meeting as listed below:

- Single Source Stares and GonoGo
- WVR Corrections
- Antenna Position Measurements
- Phase Referencing Statistics
- Astrometric Calibration and Precision
- PI Science Long Baseline Experience
- Related Higher Frequency Concerns

During the course of the meeting[1] about 30 talks were given. These talks detailed the various aspects of the test data sets, the data analysis, current implementation of software at AOS/OSF, how phases could be improved or path delays better measured, and the hardware details of the WVRs. Table 1 (Appendix B) lists the talks given in chronological order. All of the talks with presentation slide and detailed notes of each, including discussion points are available via: 'http://www.alma-allegro.nl/phase_workshop_files' (the talks and presentation notes is password protected, password: *ALMAallegro*).

The following sub-sections itemise, topic by topic, the main workshop results with relevance to ALMA observations. These report the discussions that took place. For the final recommendations see the **Actions and Recommendations** section specifically. Note, because of the mutual overlap of topics and results from various sessions, the final recommendations will be grouped with somewhat different topic headings than the outline of the workshop given above. The items below, highlighted in bold, are the more major results that need to be addressed with priority since they have either a significant degrading impact on the data and/or can improve future long baseline observations with ALMA. The case of atmospheric stability and antenna positions are also addressed in the 'Direct Impact to Observations' Section.

---

[1] http://www.alma-allegro.nl/alma-phase-correction-workshop

Single Source Stares

- Stare data must be at least 25-30 minutes long in order to examine the entire Spatial Structure Function (SSF, which is the dependence of the path length variation vs. baseline length) of the atmosphere above ALMA. This is the time that it takes for atmospheric changes to propagate over 15 km (ALMA Partnership et al., 2015, ApJ 808 1; Matsushita et al., 2016, SPIE 9906 4).

- During most times, the atmosphere SSF follows Kolmogorov (Kolmogorov 1941, DoSSR 30 301) type turbulence and is well approximated as a 'frozen-screen' (Taylor Hypothesis − Taylor 1938, Proc. R. Soc. London 164 476). There is a turn over from 3D to 2D structures around 1 km baselines, equivalent to the scale height of the 'refractive' troposphere layer (i.e. water vapour layer thickness). **In the 3D region, the observations show that the RMS phase dependence on baseline has a ∼0.6 to 0.65 slope. In the 2D regime, the SSF does not completely flatten, but still increases with a ∼0.2 slope out to >16km** (Matsushita et al., 2016, SPIE 9906 4).

- **The application of the WVR correction decreases phase fluctuations by reducing the RMS phase (over 10 sec to 120 sec) to between 25% to 50% of the RMS phase without WVR correction.** This improvement occurs for PWV >1.5-2.0 mm. In dryer conditions, the improvement factor is not so large, because there is little water vapour delay compared with other delay components.

- In a rare number of cases with exceptional conditions where the PWV is very low (<0.7 mm), very stable and the wind speed is low, the SSF has a simple 3D structure from short to long baselines. The magnitude of the phase variations are an order of magnitude below typical values (<250 μm over 30 sec). The WVRs have little effect in these conditions, as expected, since the low PWV indicates little wave vapour.

- **The short-term variability of the PWV is often associated with relatively poor phase stability, even for PWV < 1.0 mm.** Hence, a relatively small density of water vapour can be very turbulent.

- From the 2014 LBC analyses, during PWV values above 1 mm, there are correlations between the phase fluctuations and weather parameters, specifically with wind speed RMS and pressure RMS. Furthermore, this is time dependent as often higher wind speeds and phase fluctuations are seen in the afternoons. However, these correlations are not sufficiently strong to accurately derive a relation with RMS phase based upon the ∼29 data sets (Maud et al., in prep.).

- Two-Point-Deviations are a useful alternative phase variability measure compared to a simple ensemble RMS measure. Creation of Temporal SSFs shows that short baselines (<600 m) are dominated by <90 second fluctuations which are the largest in amplitude at these length scales, whereas **the longer baselines (>1000 m) have fluctuations that continue to increase in amplitude for longer and longer timescales. These are reduced by phase calibrator-target cycling every few minutes. Very 'fast cycling', < 1 min, is not effective unless the target-calibrator separation is less than about 2 to 3°** (Asaki et al., 2016, SPIE 9906 5).

- **The phase characteristics of the longer baselines of *any* array configuration are the ones which must be used to assess the phase stability for undertaking observations and to obtain good quality images.**

- The full GonoGo takes 8 minutes to check the system performance and phase fluctuations. A shorter run of only 2 min is need to determine the phase stability alone. **However, a short phase-switched observation of a quasar pair for about 2−5 min would also be useful as a direct measured of the phase errors between two sources**. These test the phase metrics and the calibrator-target angular separation and could be used to dynamically adjust the cycle times.

WVR Related Discussion

- The WVRs are excellent devices for improving the phase stability by removing 'nearly all' of the effect of the water vapour. They are more effective for the shorter term phase correction compared to longer term variations which are less effectively removed. Residual phase variations remaining are probably associated with non-water vapour components in the atmosphere.

- The Goretex 500 micron screens above the WVRs all need to be finally replaced with 25 micron FEP (Fluorinated ethylene propylene), *if not yet completed* due to signal reflections and therefore inefficient WVR correction with the Goretex screens.

- **The WVR application should be checked when their corrections are not accurate. Interpolation between antennas to fill in values for other antennas works well** – this method could also be used to obtain ACA WVR measures from the Total-Power (TP) 12 m antennas. What is the usual WVR receiver maintenance and calibration schedule?

- Leakage from the WVR receiver could cause interference or observations in Band 5. With Band 5 coming into operation, there does not currently appear to be WVR leakage near 183 GHz, but more tests are required.

- WVRGCAL and TelCal use different algorithms to derive the WVR correction, although results are insignificantly different. Improvements in the algorithms and the effects of their slightly different coupling efficiency and frequency coverage are difficult to make since the off-line WVRGCAL in CASA and the on-line WVRGCAL in TelCal have limited support.

- **The channel weights for the WVR measurement to establish path lengths are ∼fixed (brightness temperature and hence transmission dependent), and different weighting schemes have not been investigated in more detail based upon weather conditions, for example.**

- **Neither WVRGCAL in CASA or TelCal Online WVR correction include dispersive atmospheric corrections. At certain frequencies, near absorption lines, these can add up to 50 percent more path variation. Such effects will also be adverse for band to band transfer. Tests to confirm that these path-length changes can be estimated with reasonable accuracy and applied to the WVR correction are needed.**

- Tests using the offline WVR correction, suggest that a scaling factor, typically ranging from ∼0.8 to 1.5 when applied to the WVR correction often removes more of the phase variations, resulting in lower RMS phases, especially where PWV <0.8 mm (Maud et al. 2017, A&A, in press)[2]. The fluctuations corrected by scaling could be physical, due to the mixing of dry air with the water vapour and causing addition path variations. However, the scaling may also act to correct uncertainties in WVRs and coupling efficiencies, re-calibration and channel weighting schemes in the hardware and/or software. More tests are needed to confirm the application of the WVR scaling factors to the data which remain empirical at present.

- **The oxygen sounder measures the temperature profile of the troposphere above ALMA, and can possibly infer the scale height and profile of the water vapour layer. These could be used as inputs to WVR based corrections. It is unclear if this can be done in real time or retrospectively. The oxygen sounder may therefore allow improvements to the WVR corrections, and so long-term support and integration of this into the ALMA system is recommended.**

- Clouds and liquid water above each antenna produce a continuum component of the emission near the 183 GHz line. In the offline version of the WVR correction an option to fit for a continuum level as

---

[2] DOI 10.1051/0004-6361/201731197, arXiv:1707.03506

well as the water vapour line emission can be made, and is found to be effective in many cases. This option is currently best used on a case by case basis, rather than in an automatic mode, and it is not very effective for observations in very poor weather. More analysis is needed before it is included in the online WVR correction version or used automatically in the offline reduction.

- TelCal online WVR corrections produce virtually the same as that with offline WVRGCAL, but science observations still rely on the offline application since corrections associated with malfunctioning WVR systems and other experimentation of the correction are more easily made offline.
- The WVR data are sampled at 1.1 sec. It can be applied at this rate to the online data as it is collected, even if it is averaged before going to the archive. Tests show that there is very little difference in the resultant images if the data/WVRs are sampled at 1.1 sec, as compared with smoothing them both up to 6 seconds. Thus, in most science projects, averaging the data to 6 sec will not affect the images and will significantly decrease the storage required and computing time. There are exceptions however. For long-baseline data >5 km a 2 sec sampling is recommended (data and WVR), while sources that can be self-calibrated should be sampled at the one-second level as well as radio emission sources that vary on one second times-scales. Projects with such constraints may require particular attention in the initial setup stage.
- The antenna positions are determined by all-sky observations of quasars (more details below). The online WVR corrections use the full wet path option and are applied in the TelCal software which then determines the antenna positions. For consistency, the WVR corrections, either done online or offline must be applied for all the data, whether for testing or for PI experiments.
- **Using only one weather station for WVR correction is an over simplification. Using all weather stations needs to be tested and could give more accurate WVR corrections, as would a better temperature profile that could be determined from the Oxygen sounder (see above), or a diurnal model based other data.**
- There is an unsolved issue in TelCal with how coefficients and ‘Tref’ are used by the correlator, and moreover a problem that Sky coupling efficiencies in some cases were slightly larger than 1. Further calibration of these data is recommended.
- Only in a few datasets, and only for subsets of baselines with a weather station near each antenna, are there time-matched phase variations correlated with pressure variations. The correlations found in 2-3 datasets are over timescales <5 min and possibly trace the dry air component. Shorter time variations are not correlated. Such correlations may occur in dry conditions where the WVRs are less efficient. Even though there is an occasional correlation, the scaling of the pressure variations to phase variations are not understood. These results suggest a weather station located near each antenna will **not** be effective in removing the dry-air phase changes.

Antenna Position Measurements

- In the 2014 LBC the antenna positions, as measured by the all-sky-delay calibration, varied considerably (order of millimetres). Using full wet path option option in TelCal, the antenna positions were somewhat more consistent during the 2015 LB observations (within a few millimetres).
- Changes in antenna position on longer baselines can be as large as *5 mm* (per all-sky-delay run), and is significantly in excess of the design specification of 0.050 mm which is unreasonable for the long baselines. These changes are mostly in the vertical direction, not the north and east ground displacement of the antenna position; hence are likely caused by the change of antenna delay. These changes produce large imaging errors and can affect the bandpass slope for the target.
- For the 19 high-quality all sky runs, the dispersion in the Z (vertical) direction is 0.2 mm per km. This will produce phase errors which scale with the angle between sources on the sky, so a calibrator within

$< 3^\circ$ of the target is recommended. At band 3, a suitable calibrator for most targets can be found, but may need confirming short observations (weak calibrator survey or cone-searches) for confirmation. For higher bands, calibration using the band-to-band, e.g. band3-to-band7, calibration scheme to use a calibrator closer to the target would be very beneficial. This recommendation is discussed later.

- Splitting the all-sky delay runs into three-15 minute runs and calculating the change in the vertical component of the baselines for the longest spacings, finds that this component is stable to about 1.1mm over an hour. Thus, Mini 'all-sky' runs of 15 minutes could be taken before (and after) PI LB observations to obtain the apparent antenna offset in the vertical direction.

- There is little correlation between the pressure differences among the various weather stations and the measured antenna position offset, even after correction for their height differences. However one night, during which there were continuous baseline runs, showed some correlation between pressure differences over the night and apparent antenna position changes. The antenna position 'changes' of the longer baselines were much larger than that expected given the pressure changes according to theoretical predictions.

- It is possible that the cause of some vertical position errors are due to inaccurate WVR corrections, perhaps due to coupling efficiencies, less than optimal calibration or the assumed troposphere model, e.g. scale height and lapse rates for the atmospheric model used. Additional tests are being planned - see Actions and Recommendations.

- The system of keeping track of the ALMA antenna positions in the TMCDB and results from the antenna-position measurements (all-sky-delays) is a bit cumbersome. The distinction between antenna and pad positions is arbitrary and they should be merged into one table.

- It is unclear when antenna positions are not updated after an all-sky delay observation: because the position change is $<50\,\mu$m or less than 3σ?; or is it because the antenna was not included in the observation. Thus it is difficult to understand if there are systematic changes if 'small' ones are not recorded.

- Some antennas showed systematic slow position changes of 0.5 mm over a four month period in 2015. This is anomalous and not understood, but is the *only* instance of such a change. This effect suggests that there may be real changes in the antenna positions caused by a pad settling or by motion of the region where these antennas are located. Volcanic activity can show vertical movements of $10-20$ mm per year, and may be related to these systematic antenna position drifts.

Phase Referencing Statistics

- The 'weak calibrator' survey program is continuing as a low priority filler program. Over 2000 sources >80 mJy at 20 GHz have been added from the AT20G and VLA catalogs and the ALMA flux density at 100 and 290 GHz will be measured. Some of these weaker sources do not have VLBI positions, so have apriori position offsets as large as $1''$ which is too large for most ALMA imaging needs. Additional ALMA observation may be needed to determine an improved position to $0.01''$ as well as continued flux monitoring.

- There can be differences in VLBI positions, most of which are measured at 8 and 20 GHz, and the ALMA positions, usually measured at 100 GHz. Some differences are obvious from the ALMA images that show two components separated by $1''$, one of which is probably the VLBI component. These sources are not used as ALMA phase calibrators. More insidious are those quasars with strong cores with a peak position that can change with frequency. This apparent motion is called 'core-shift' and is associated with the change of position caused by the synchrotron optical depth dependence with frequency. Such position changes between 8 GHz and 100 GHz is at the $\sim$3 mas level, and are usually less than the accuracy needed for most ALMA high resolution images.

- Multi-frequency calibrator surveys with the KVN (Korean VLBI Network) could be useful for ALMA to determine the spectral index of quasars and perhaps measure the apparent position shift with frequency. Discussions between the two observations are continuing.

- Obtaining calibrator data from the archives from groups outside of JAO is not straight-forward. Permissions for their use must be made more streamlined.

- **Fast switching with 20 seconds cycle times results in much lower phase variations after calibration than at a comparison 2 minute time scale if the phase calibrator is close to the target <3 degrees. However, for phase calibrator to target separations of >3 degrees the 20 second referencing timescale phase variations start to become indistinguishable from those for a 2 minute cycle time. This is because at larger source separations the different atmospheric variations between the two sources are significantly different, even at the same time** (Asaki et al., 2016, SPIE 9906 5).

- After fast switching phase referencing often there is still an increase of phase variations with baseline length in Spatial-Structure-Function (SSF) plots that should have been calibrated out. This difference appears to scale with source separation angle. Two explanations are that a high atmospheric layer of turbulent 'dry' air is responsible, or a thick layer with large scale height. A 'dry' component may be able to explain the SSF shape of the QSO stares taken in very dry weather. In these cases the slope increase is constant (baseline$^{\sim 0.65}$) associated with only a 3D atmosphere structure without the breaks according to Kolmogorov type turbulence (Kolmogorov 1941, DoSSR 30 301, see also Single Source Stares Section).

- **The current GonoGo is a single-source observation to calculate the phase RMS over 2 minutes. These are not phase referencing observation (i.e. with 2 sources separated on the sky) and therefore the variability measured does not include those incurred when switching sky position between the calibrator and target. Furthermore, the ∼30° GonoGo phase RMS limit (at the observing frequency) translates to a coherence loss of about 13% over 2 mins. $\mathrm{T}^{0.5}$ is used to account for RMS changes based on various ALMA default cycle times per band and baseline configuration at present. However, it should be possible to implement a method to accurately measure the time-dependence from an improved GonoGo (see recommendations).**

- Phase referencing becomes increasing harder at the higher bands, particularly 8, 9 and 10. Generally, the RMS phase stability, as measured by the GonoGo, must be lower than 30° at the observing frequency, and a sufficient bright calibrator must be detected with >10 SNR per antenna solution within about 10° separation from the target. For longer baselines closer calibrators must be used in order to phase reference faster and improve coherence. Currently up to 120 sec is required to attain the SNR for in-band calibrators which can lead to problems with accurate amplitude calibration due to poor phase-up.

- **According to GonoGo statistics bands 9 and 10 can sometimes be undertaken until ∼9 am Chilean Local Time.** However, the pointing and focus at these bands is also more variable during the day.

- Cycle 5 limits (as of the meeting date - May 2016) the baseline lengths as a function of frequency to: ∼15 km for bands 3-4-5-6; ∼10 km for band 7; 1.5 km for bands 8-9-10. In terms of the increase of short-term phase variations with baseline-length, measured from from the atmospheric SSF results, there is little increase in phase RMS beyond 1 to 2 km. Hence, the GonoGo phase variations at 2 km are usually not much less than that at 15 km. Based *only* on short term phase variability the longest baselines can be used at higher frequencies if the strict calibrator criteria is met (short cycle time, strong and close calibrator). These parameters cannot be met with current in-band calibration as calibrators are weak and distant (>6°). Systematic phase difference are present between the calibrator and target, in part caused by antenna position errors which increase with baseline length and moreover the calibrator-target separations. As the phase discrepancies increase with frequency the higher frequencies should be limited to baselines of ~1.5 km in Cycle 5. Alternative calibration using the band-to-band technique requires a specific study if the longest baselines are to be successfully used at bands 8, 9 and 10.

- If the target can be self-calibrated, then the nominal GonoGo phase metric limitation can be relaxed. The stronger the long-baseline emission from the target, the shorter the self-cal solution interval that can be made, in some cases down to a few seconds. If the self-cal interval is longer than about 15 sec, the very short-term variability cannot be removed. However, longer solution intervals up to 20 min for relatively weaker sources or those somewhat resolved at longer baselines are still useful for removing baseline errors and large-scale atmospheric errors.

- The decision (as of May 2016) was made for Cycle 5 *not* to support longer (>1.5 km) baseline projects at bands 8,9,10 even if they could be imaged with self-calibration. Including self-calibration in the current scheme with the Observing-Tool (OT) to estimate on-source time and scheduling would require specific attention.

- In very poor conditions fast switching can also help, provided you can still detect the calibrator source which *must be within 3 degrees.* **A more dynamic way to choose cycle times would allow the best strategy to be implemented given the conditions and calibrator separation - this would guarantee the final coherence quality of the observations**.

- The SMA attempted to use the Ozone lines in the astronomical receivers for phase correction. Using the Ozone as a back light through the water vapour screen. The tests showed this was possible, but took too long to be feasible for PI observations. This could be tested at ALMA, along with possible measurements of Oxygen lines to investigate the dry term fluctuations. Variations in line profiles could be easily modelled with existing codes, e.g. ATM model, for improved phase correction.

- For weaker calibrators, it may be possible to bootstrap out the calibrations from shorter spacings to longer spacings. Also, the grouping of close antennas would help in obtaining more signal-to-noise for phase solutions. Software for this requires some development.

Astrometric Calibration

- **The nominal positional accuracy of a target image is about 1/20 of the synthesized beam, with a Gonogo phase RMS of about 20° (somewhat better than the minimum requirements and with a phase calibrator within 3 degrees of the target). This assumes that the image RMS of the target is $> 20:1$, otherwise, the positional accuracy will be based on the target SNR limits. This estimated astrometric error scales roughly with the target-cal separation and the phase RMS.** However, there is a minimum positional accuracy of about $0.01''$, regardless of the target strength and observation resolution, if only one phase calibrator is used.

- Multi-source calibrations are recommended to improve the above astrometric accuracy if needed. The reduction methods are non-standard, but the simplest case of a target in line between two calibrators (within about 20° linearity), can be handled easily.

- The general astrometry observations require three calibrator sources around the source, within about 10°, which may not be possible for many targets.

- The variation of the positional accuracy with configuration size is **not** linear, but varies as $B^{-0.5}$ where B is the maximum baseline length. Also, the positional accuracy is formally independent of frequency (see next bullet) because both the resolution and systematic errors scale with frequency.

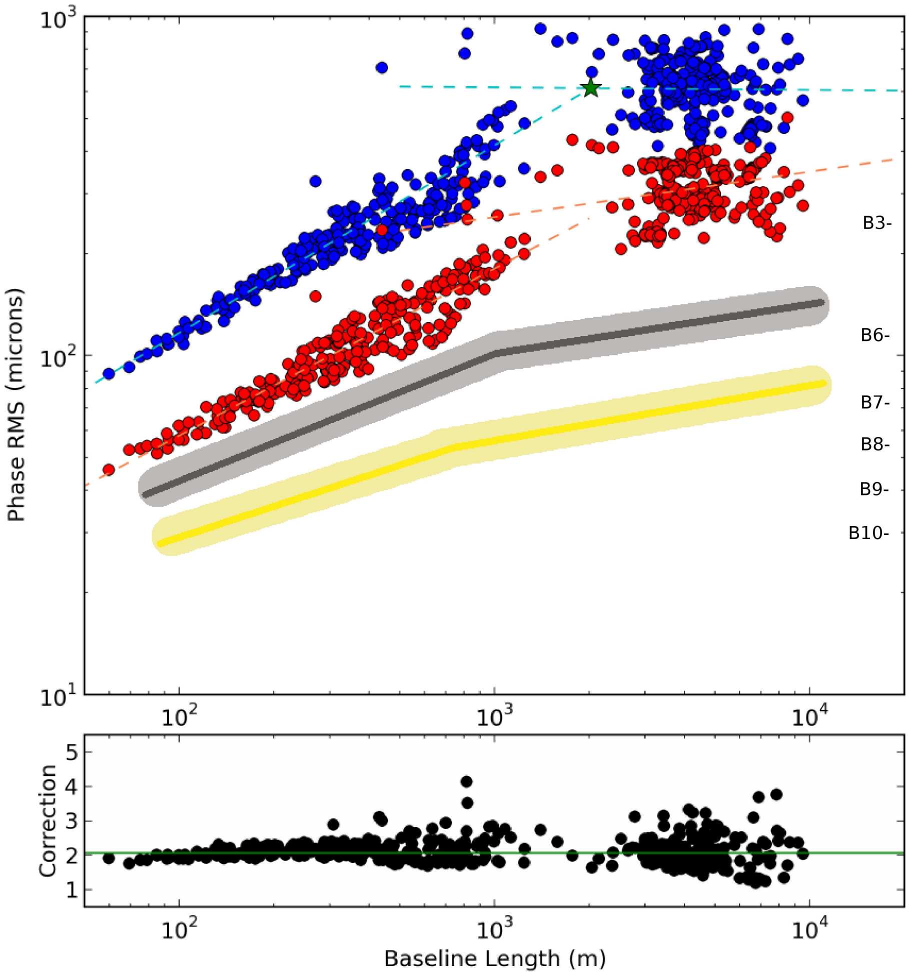

Figure 1: *The Spatial Structure Function (SSF) of a QSO stare taken in reasonable low PWV conditions ~0.6 mm. It is representative of a typical SSF. Blue and red circles are the phase RMS in microns for each baseline before (raw data) and after WVR correction respectively. The plot at the bottom shows the WVR correction factor as a function of baseline length. (factor of ~2) The grey and yellow regions indicate the reduction in phase RMS due to fast switching (to a calibrator <3 degrees away) on timescales of 2 minutes and 20 seconds respectively. For more distant calibrators (>5 degrees) switching faster does not generally improve the phase RMS past that shown at the 2 minute level. The regions are highlighted with a more solid colour to show the break, or turnover, to a shallower SSF occurs at shorter baselines for faster switching. The RMS values are approximately constant after baselines of ~1-2 km, hence there is no need to limit baseline length based on phase stability arguments. If one can calibrate 2 km baselines, then one can also calibrate 10-15 km baselines. The SSF from a stare does not account for systematic baseline length based errors, e.g. antenna positions, or for the target-calibrator separation. Note that the phase RMS, even after Fast Switching does not drop as low as ~30 μm as required for 30 degrees phase RMS at Band 10. The various 30° phase RMS levels are indicated to the right of the plot and are ~250, 110, 70, 55, 40, 30 microns for bands 3, 6, 7, 8, 9 and 10.*

- For thermally-emitting objects, ALMA Band 7 is usually the optimum frequency since the ratio of flux-density to system temperature is usually at a maximum. However, the configuration chosen should be one where the target is not significantly resolved.
- A detailed astrometry experiment was made during the 2015 LB observations. Four bright quasars were observed alternately over many one hour experiments at band 3, with each scan about one minute

in length. The sources were within a radius of 10°. **From the statistics over two months of ten such observations, the following ALMA astrometric accuracies were obtained: For a target-cal separation of ∼5°, the RMS position error of the target was 4 mas. Using two calibrators on either side of a target, the RMS position accuracy was decreased to 0.5 mas.** These results can be scaled, with baseline length, as described above.

- Accurate parallax determinations require observations at four specific periods during a year (need not be in the same year) in order to decouple the parallax and proper motion of the target. Obtaining such observations with similar resolutions are difficult to obtain with the present ALMA configuration schemes. If only the proper motion is needed, then observations at any time of year are fine. Of course, the proper motion accuracy varies as the span of time of the observations.
- Check sources are used for all long baseline and all high frequency PI observations. These sources are often too weak in order to use them as a secondary calibrator since antenna-based phase solutions are needed for every scan. However, the image position compared with their known position and the image quality will give a realistic estimate of the astrometry and image quality of the target.
- A better, more flexible, astrometric mode is needed for the OT generation if PIs required such accuracy. In most cases, the PI would need to choose the calibrators and some weak calibration searches (see below) may have to be made to find the optimum configuration of calibrators for the target.

PI Science Long Baseline Experience

- There were some hardware and software problems with longer baseline observations. These included delay jumps and delay server timing errors due to the large number of antennas which possibly overtaxed the system when short integration times are used.
- **There is an issue with residual phase slopes in the phase calibrator and target source, after calibration using the bandpass source.** It is likely that these residual phase slopes are associated with the large antenna position offsets (mostly vertical component). With the typical angular separation between the bandpass calibrator and the target/phase calibration, some residual phase slopes are consistent with the antenna position errors.
- There are no statistics of the properties of the check sources for the delivered LB projects; for example, position offset, image quality, correlation of check source quality and target quality. These *should* be gathered in order to assess the future use of the check sources.
- **It is most convenient to use an unchanging reference antenna, especially during the long baseline observations, in order to register the (apparent) antenna positions as accurately as possible.**
- There were hardware problems due to high data rates even after significant averaging of the data. **With even more antennas, this could be a problem for Cycle 4 and future Cycles. Data averaging to 2-sec or longer could alleviate the problem, but it must be investigated soon as the 5 km and longer baselines are scheduled.**
- There were a significant number of long baseline projects observed in such bad conditions and could not pass QA2, it is unclear whether the GonoGo was overlooked as some scan to scan phases were offset >100°. **One of the most important improvements before the next LB observations is to improve the TelCal and AQUA software used during and just after the observations to assess the scan to scan phase changes of an experiment. This is an important QA0 assessment.** Additionally short target-cal pre-observation tests may be required to see if the science observation is feasible as well as an improved GonoGo indication of conditions.

- Self calibration was used for those long baseline projects for which there was sufficient flux density at the longer baselines. However, the normal phase referencing *must* be sufficiently accurate in order to obtain a reasonable quality image for the first self-calibration round. For weak sources that cannot be self-calibrated, much better than average phase stability is needed (see above on feasibility test).
- **There is uncertainty in ALMA about how to deal with self-calibration; where it belongs in the proposal procedure, the OT options and data reduction process. These discussions should be continued and emphasis on achieving the good quality images. How much should be done by ALMA and how much by the PI is also under discussion.**
- Band 7 data was successful with >10 km in 2015. For this reason Band 7 proposals will be accepted for Cycle 5 with baselines up to 10 km. Band-to-band calibrations will be very important in using a phase calibrator close to the target. This technique will be investigated throughly in Cycle 4/5 and perhaps included in future cycles.

Higher Frequency Phase Referencing

- For phase referencing at bands 8,9 and 10, the average distance between a detectable phase calibrator (assuming 8 GHz bandwidth) and a random target is 4°, 7° and 10°, respectively. These relatively large separations produce systematic phase errors between the calibrator and target that produce non-negligible image distortions. Calibration techniques such as band-to-band need more investigation in order to allow the use of much closer calibrators.
- In order to find the closest detectable calibrator to a target, especially those at bands 7 or higher, an **Allegro-ALMA initiative, called cone-searches** has been in use for many months. There are thousands of cataloged radio source, mostly from the AT20G and VLA catalogs, that could be strong enough at ALMA frequencies, but yet do not have even rough estimates of their flux density and angular size at frequencies above 30 GHz. In order to find suitable calibrators around the PI targets that will be observed in the near future, short observations at bands 3 and 7 (and 9 where possible) are made in order to check their viability. Results from several months of cone-searches show that even at band 10 a calibrator within $\sim 10^\circ$ of a target can be found for many PI targets.
- An alternative phase referencing method called **band-to-band calibration** uses phase calibrator observations at a lower frequency to apply to a target observation at a higher frequency. This transfer assumes that the temporal phase changes are non-dispersive (delay-like) and scale with frequency. In addition there is an instrumental phase 'offset' between the two bands that must be removed.
- The band-to-band instrumental antenna-based phase offset is measured using a strong source, called the Differential Gain Calibrator (DGC). The instrumental offset is constant throughout the observations, but because the phases can significantly vary—almost entirely from the atmosphere component especially at long baselines and at high frequencies—the DGC observations must alternate as fast as possible (20 s cycling) between the high and low frequencies in order to minimise the effect of these changes.
- The phase difference between bands may not only depend only on their instrumental (internal) phase differences, but also on a another component that depends on the position of a source in the sky. For example, if two DGCs could be observed at the same time but at a different sky position, then their phase offsets could be somewhat different. If the array is well-calibrated and there were no underlying atmospheric phase variations, there would be no difference between DGC sources. **DGC observation schemes to investigate this additional source to source phase difference are now being tested - implementing the rapid cycling (20 s) to understand if the differences in offsets previously found were largely atmospheric.**
- Dispersion of the radio signal near strong atmospheric lines produce non-frequency-scaling phases that cause errors in any band-to-band scaling. These dispersion effects can be estimated from the PWV and

frequency and removed from the observed phase, although these methods have not been implemented yet.

- For long baseline observations at band 7, Band 3-to-Band 7 fast switching calibration may be useful in order to use a calibrator at band 3 that is within 3° the target. Such experiments are also being tested (and other frequency pairs), and a standardized mode for the OT and CASA based calibration must be developed.
- Tests to approximate a simultaneous dual frequency array were made by making two sub-arrays in which an antenna in each subarray was paired with one in the other in the sense that their separation was as close as possible. Observations of one quasar with the two subarrays show that the phase at the lower frequency can be multiplied up to the higher frequency. This technique works best in the smaller configuration when paired antennas can be separated by less than about 50 m. The number of baselines for the observations at both frequencies was about 1/4 of that if using the entire array. A minor problem is that that the CASA based application of frequency-paired antenna-based gain tables requires some reordering of the gain solutions to account for different antenna index ordering.

Impact of Phase Variations to Observations

Many of the topics, mostly on various properties of the phase stability, that were discussed at the meeting have a direct impact on observations at ALMA. Some of these have a dedicated discussion below. The instrumentation and tests for understanding the phases that were discussed in the meeting have already been reported above.

Before listing the actions, recommendations, and further analysis and tests, the three main components of the phase variations and properties are given below. The first two properties can be inferred from single source observations, the third only with phase referencing tests.

Short term phase variations: The short term phase variations between phase calibrator scans affects the coherence. The level of these depends on the cycling time (i.e. consecutive observations of a phase calibrator - ranging from tens of seconds to the maximal ALMA scan which is usually no longer than about 5 min), weather/stability conditions, frequency and baseline length. This is the statistic that is roughly measured by the GonoGo short observation that is run several times a night or when the phase stability appears to be changing. The short-term phase RMS of 30° results in a ∼13% decoherence, where coherence follows exp($-\phi^2_{\rm rms}/2$), this is generally the GonoGo limit for scientific observations at a specified frequency - although timescales related to cycle times are **not** directly measured in all cases. A more relaxed phase RMS of 45° drops the coherence to 73%, and may be required for higher bands given weather conditions rarely meet the 30° limit. Since the phase fluctuations scale with frequency, the results of the GonoGo observation at any frequency can be scaled to the observations of the pending scientific observation. The current test calculates a fixed 2 min RMS value but more accurate time-dependence, based on cycle times could also be implemented in a future version according to scaling found from the QSO stare data.

Scan to scan phase changes: The fundamental calibration associated with phase referencing is the interpolation of the measured antenna-based phase between successive calibrator scans, typically from 1 to 5 min apart, to the target. The phase change between scans should not be larger than about 100° for accurate interpolation, and even this large a value suggests that the phase associated with the target may be significantly different to that interpolated between phase scans. Currently it is thought that a 60° change between scans at the longer baselines should be the limit for successful phase referencing. This scan-to-scan phase metric is related to the short-term phase RMS and the spatial structure function. The initial tests show the 2 min scan-to-scan phase is approximately double the short term phase RMS measured over a 2 min interval (as per the GonoGo).

Angular Structure of the Phase: In addition to the short-term and medium-term phase variations for any one source, there is also a dependence of the antenna-based phase with angle in the sky. This dependence is caused primarily by: array-parameter errors such as antenna position error; and large-scale atmosphere structure, mostly in the dry component, since the WVR receivers remove most of the wet delay component when it is significant. As the quasar-pair phase referencing tests have shown, this phase error component varies approximately linearly with calibrator-target separation, and is somewhat **decoupled** from the short-term phase changes which are associated with small-sized water vapour turbulence. The amount of coupling depends on the phase cal cycle time and the target-calibrator angular separation and has been discussed previously. Generally, a calibrator within 3° is recommended, but this is not always possible for frequencies above 300 GHz.

Actions and Recommendations

Here we amalgamate the previous results into a coherent set of actions and recommendation looking ahead to 2016/2017 PI observations and LBC testing. The sub-sections are arranged such to emphasise the recommendations for testing, or PI science observations, but also to include actions to be taken for data in hand, and checks of the hardware and software at ALMA that have an impact on observations. Furthermore, a section is included specifically for future testing ideas that could improve observations at ALMA. Note, although we have categorised the recommendations into the most relevant sections some do overlap with more than one category. The recommendation items are enumerated for ease of reference in any future work or correspondence.

A) Follow-up Analysis of Existing Data and Software
Ordered from higher to lower priority.

1. Further analyses of the existing multi-separation QSO observations to confirm the switching/cycling improvements as a function of QSO separation, and weather conditions.

2. Test and implement a new GonoGo system consisting of a 2-3 minute observation with fast switching on QSO pairs to understand the atmospheric fluctuations as a function of time, baseline length and source separation. Derive the conversion factor between GonoGo phase switching tests and the expected phase RMS given the cycle time and source separation.

3. Establish the viability of a WVR-correction scaling factor and any correlation with meteorological properties (Maud et al., 2017, A&A).

4. Improve the algorithm and observation scheme for the interpretation of the GonoGo tests, based on the quasar stares and quasar pair phase referencing. Can cycle times also be estimated?

5. Investigate antenna-combination schemes for self-calibration of extended sources, or determining the phase of weak calibrators. This could include modelling a rolling screen over the array.

6. Establish whether the PI science bandpass observation can ultimately replace the GonoGo assessment and be used to determine any WVR-correction scaling for the observation − implemented as of July 2016.

7. Further examination to see if the long term pressure changes are correlated to antenna position changes.

8. Further examination to see if the short-term pressure changes are correlated with antenna phase changes, or other meteorological correlations, such as wind speed, RMS changes of PWV and wind. This can only be done with antennas close to a weather station. The results so far show little correlation.

9. Investigate how the adjustment of atmospheric parameters affects WVR phase solutions, in TelCal offline or WVRGCAL.

10. Use of the Oxygen sounder outputs to retrospectively adjust WVR corrections. TelCal (offline) and WVRGCAL must be edited, although this requires expert knowledge of the codes. The oxygen sounder should be integrated into the ALMA systems to provided correct time-matched meteo-data.
11. Examine results for Oxygen sounder and stare data if taken at the same time to better understand the atmospheric conditions, especially the extraordinarily excellent, low phase RMS nights.
12. Investigate the use of Oxygen sounder temperature profiles in ATM cal. See how the temperatures change compare to the pre-set winter/summer lapse rates now used. Look into using standard diurnal profiles for the given time of year, or directly the measured profiles.

B) WVR Improvements

1. Investigate and clarify of how TelCal passes 'Tref' and coefficients to the correlator for WVR correction.
2. Investigate application of a 'full wet path' for science data.
3. Understand why sky coupling values >1 and get the agreed WVR efficiencies for TelCal in WVR correction.
4. Include dispersive path changes in TelCal and WVRGCAL (if not already implemented correctly).
5. Implement the cloud removal algorithm, correct WVR efficiencies and WVR channel frequencies in TelCal *and* WVRGCAL. Test first offline then implement into online Telcal.
6. Use all weather stations for WVR relevant corrections and quantify the changes.
7. Understand if WVRGCAL can be modified, not just maintained, or if a different offline WVR correction software is required (i.e. TelCal - offline). This requires expert knowledge of these codes.

C) Recommendations for Cycle 4-5 LB Testing

These recommendations are in roughly priority order, from higher to lower.

1. Check long baseline robustness using as many antennas as possible, high data rate and FDM observations. Look for delay jumps, delay zeros, visibility amplitude zeros and other issues.
2. Test the accuracy and applicability of including a 15 min mini-Zenith-Delay run before most LB (>5km) to remove systematic delay errors. If useful, develop software for its application.
3. Higher bands may need band-to-band calibrations and should be tested, first at bands 3-7 to examine the best implementation and DGC limitations.
4. Test how band-to-band is affected by dispersive effects. This can be done at shorter baselines. Can the dispersive phases be estimated from the frequency and PWV only?
5. Investigate effects on amplitude of band-to-band phase referencing. How will observations do time dependent amplitude calibration, are there intrinsic offsets between bands? This can be tested at shorter baselines.
6. The dual-frequency sub-array tests should continue and determine if the relatively large separation between pair antennas is a problem.
7. Observe in Band 3 to investigate the use of the Oxygen line for correcting atmospheric fluctuations.
8. Have more antennas on longer baselines pads to better examine if antenna position changes are local or systematic.
9. Test use of measured Ozone lines acting as a back light through the water vapour layer to provide a wet path correction and compare with WVR corrections.

10. Push long baseline high band (8, 9 and 10) tests, both in-band and band-to-band. If these bands can be done at 2 km, then according to the phase metrics they can be done for configurations with baselines >5 km if calibrators meet the criteria.
11. Complete the change of all Goretex WVR screens to FEP.

D) Recommendations for PI Science observations, emphasizing long baselines

1. The phase calibrator−target separation should be as small as possible, and combining SPW and polarization to detect weaker calibrators is recommended (further investigations are underway for best implementation). For configurations with >5 km baselines (currently bands 3,4,5,6,7), three degree separation is needed unless self-calibration is *assured*.
2. The LB antenna positions must be measured with very good phase stability (current GonoGo at 10 km of < 15°, generally with pwv <1.0 mm) as soon as possible, although PI observations can be done before this all-sky-delay. Repeat every two weeks, if possible.
3. Decide on a reference antenna before the long baseline configuration. Generally, one antenna will remain on a pad near the centre for about 6 month, for example DA41 was fixed from Dec 2015 to July 2016.
4. Calibrator Cone searches and weak calibrator survey observations should be made for long-baseline calibrators, if not already known, well before the LB observations in order to be done with relatively short baselines (better phase stability and less use of LB time).
5. QA0/AQUA analyses of LB projects (and high frequency projects) must be made just after completion of the observations. One of the most important parameters is the average phase difference between calibrator scans. If it exceeds about 80° for the longer baselines, the image quality may be poor. Generally, a GonoGo condition of < 30° (current test) will provide the above scan to scan limit.
6. We recommend 6-sec, 3-sec, 2-sec integration times for configurations < 2km, 2 to 7 km, and > 7km, respectively. If the target can be self-calibration at a short-time scale or has very variable emission (e.g. the sun or pulsars), shorter integration times are needed (∼1 s).
7. Offline WVR correction should be used for all science reductions. The online 1 sec WVR corrections are needed for the QA0, GonoGo and TelCal applications during the observations. For the archive, *only* the WVR-uncorrected data are needed, but at the present time both corrected and uncorrected data are stored (i.e corrected is with the online-TelCal application).
8. Most science LB data in Cycle 4 will not be standard and will be processed with manual reduction and experienced data analysts should be used.

E) Future Tests for Site or Atmospheric Understanding

1. Use accurate GPS equipment to monitor if antenna position changes are caused by physical site changes. ESO contacts should be able to facilitate the hire of such equipment.
2. Explore the use of InSAR (Interferometric Synthetic Aperture Radar) to study the deformation of Pyroclastic shells/flows on Chajnator. Contact Volcanologist/Geologist for possible interests in site studies.
3. Use the ALMA antennas at the water line (band 5) to do atmospheric tomography by focusing on water vapour layers about 2 km above the ground. Do these have any correlation with the WVR measured. Such data could improve the atmospheric model parameters.
4. Use available models to examine site deformation based on gravitational loads, e.g. Tides. Some of these effects are already in the Calc software used for the online delay model, but may not be turned on.

5. Investigate atmospheric models with dry and wet air mixing (c.f. ALMA Memo 517) to test if this explains the WVR scaling factor.
6. Explore potential use of LiDAR and/or SODAR for measuring atmospheric structures, and making terrain models if airborne. Doppler LiDAR, and SODAR can both be used for wind speed measurements and compared to ground weather stations. LiDAR can also act as a remote instrument to measure densities of atmospheric constituents, and could be used for dry term corrections.

Conclusions

We have presented a summary of the May-June ALMA-Allegro phase workshop concerning the impact that atmospheric instability, particularly the variable and complicated delay properties over the array. We have described the relevant crucial ALMA instrumental features, summarised the many tests and their results and reviewed the long baseline PI observations in 2014 and 2015. We also included some related high frequency calibration concerns.

We then suggested recommendations for additional analysis of existing data with hardware, WVR improvements, additional LB testing including good support of PI observations in 2016-2017 and more long term analysis. Some of which are already being tested and have been implemented as of the date of this report.

Finally, an ALMA-phase working group will be created in order to improve communication of observational plans and tests underway, and to hold regular telecons to update all members working directly on data.

APPENDIX A

Table 1: *List of Participants.* * *= attended via video-con.*

Name	Email
Yoshiharu Asaki	yasaki@alma.cl
George Bendo	george.bendo@manchester.ac.uk
Michael Bremer	bremer@iram.fr
Crystal Brogan	cbrogan@nrao.edu *
Dominique Broguire	broguier@iram.fr
Yanett Contreras	ycontreras@strw.leidenuniv.nl
Bill Dent	wdent@alma.cl
Ed Fomalont	efomalon@nrao.edu
Ciriaco Goddi	cgoddi@strw.leidenuniv.nl
Antonio Hales	ahales@alma.cl
Richard Hills	richard@mrao.cam.ac.uk
Michiel Hogerheijde	michiel@strw.leidenuniv.nl
Todd Hunter	thunter@nrao.edu *
Taehyun Jung	thjung@kasi.re.kr
Satoki Matsushita	satoki@asiaa.sinica.edu.tw
Luke Maud	maud@strw.leidenuniv.nl
Bojan Nicolic	b.nikolic@mrao.cam.ac.uk *
Scott Paine	spaine@cfa.harvard.edu
Juan Pardo	jr.pardo@csic.es
Neil Phillips	nphillip@alma.cl
Anita Richards	a.m.s.richards@manchester.ac.uk
Ian Stuart	ims@strw.leidenuniv.nl
Remo Tilanus	rtilanus@strw.leidenuniv.nl
Carmen Toribio	toribio@strw.leidenuniv.nl
Catherine Vlahakis	cvlahakis@nrao.edu *

APPENDIX B

Table 2: *Workshop talks presented at the meeting.*

Date & Topic	Talk Title (*) discussion rather than presentation	Speaker and Presentation
Monday 30 May	Meeting Outline and Expectations	Ed Fomalont
Single Source Stares	(continuous observation on one source)	
	Phase Characteristics of ALMA Long Baselines	Satoki Matsushita
	Phase Metrics Analysis & Weather Correlations, Two-Point-Deviations	Luke Maud
	Go/noGo overview	Bill Dent
Tuesday 31 May	WVR Related Discussion	
	Introduction to Radiometer-based Phase Correction	Richard Hills
	The Oxygen Sounder	George Bendo
	Optimising the WVR Solutions with Scaling	Luke Maud
	Cloud Continuum Removal (*)	Bill Dent
	TelCal online WVR	Dominique Broguier
	Pressure and Phase Comparisons	Satoki Matsushita
	WVR Smoothing from 1 to 6 Seconds	Ed Fomalont
	PdBI/NOEMA Phase Correction	Michael Bremer
Antenna Positions		
	Inside TelCal Antpos	Dominique Broguier
	Antenna Position and Weather Station Measurements	Todd Hunter
	Antenna Position Updates	Antonio Hales
	Antenna Position Changes and Zenith Path Delay	Ed Fomalont
Wednesday 1 June	Phase Referencing Statistics	
	What about the Calibrators	Anita Richards
	Phase Referencing Statistics and Correlations & Fast Switching Experiments	Yoshiharu Asaki
	Is GonoGo good for understanding Phase Referencing (*)	Bill Dent
	KVN MASK multi-frequency Calibrator Survey	Taehyun Jung
	Pushing to Higher Frequencies at Longer Baselines	Ed Fomalont
	Does Very Fast Switching Help in Marginal Conditions (*)	Yoshiharu Asaki
	Probing the Troposphere	Scott Paine
	Properties of LBC statistics from GonoGo	Bill Dent
Astrometry		
	Astrometric Calibration	Antonio Hales
	Astrometry Considerations	Ed Fomalont
	Check Sources in Astrometry (*)	Ed Fomalont
Thursday 2 June	ALMA Long Baseline Experience	
	Thoughts from the Cycle 3 Long Baseline Campaign	Crystal Brogan
	Check Sources for PI Observations	Todd Hunter
	Experience from the LBC and Band-Width Switching	Anita Richards
Higher Frequencies		
	Band-To-Band	Ed Fomalont
	Band 3 to Band 7 Fast Switching	Yoshiharu Asaki
	Simulating Dual Frequencies with Sub Arrays	Luke Maud
	High Frequency Calibrators	Yanett Contreras & Carmen Toribio
Friday 3 June	Open Discussion Morning	All Participants