VLBA Scientific Memo #37

Flux Density Calibration on the VLBA

R. Craig Walker
Dec 15, 2014

Abstract:

Reports from the MOJAVE project and from our own science project BW102 indicated that there were errors of 25% to 30% in the flux densities measured on the VLBA with data from the PFB personality of the RDBE. This memo details what was learned while trying to understand these errors. A test observation was made on the VLBA plus the phased VLA. The internal VLA data, which included 3C286, were used to determine the flux densities of the VLBI targets. The whole amplitude calibration process was reconsidered carefully, resulting in a number of suggested changes to both procedure and software. The details of what was found, and the new recommended procedure, are the subjects of this memo.

The new procedure involves five steps:
1) Scale the data to give unity average autocorrelation across the full band (ACCOR). This step has not changed.
2) Remove residual delay offsets from a reference scan (PCCOR or FRING) to prepare to make a bandpass. This used to be done later, but the procedure has not changed.
3) Form a bandpass using the full band and power normalization (BPASS). This should be done on cross correlation data on a strong source to get the phase bandpass.
4) Scale the data by the small factor required to make the calibrated autocorrelation values unity in the channels to be kept (new task ACSCL). This step is required because the normalization of autocorrelation and cross correlation based bandpasses differs slightly because of non-correlating, aliased data in edge channels of the autocorrelations.
5) Apply the Tsys and gain data (APCAL). Unwarranted fluctuations in the Tcal values with frequency need to be dealt with. Opacity corrections should always be made. The gains used should be those derived from the RDBE and should be based on the Perley and Butler (2013) flux density scale.

The above prescription results in VLBI flux densities in the test observation that were are about 4% below the VLA values. The match for BW102 is not as good, with an offset of 8% to 17% depending on exactly what is being compared. The reasons for these residual offsets in the older data are not clear.

1: Introduction

In June of 2014, Yuri Kovalev, representing the MOJAVE project, called our attention to a large change in the flux density calibration results in their 2cm data after they started using the RDBE. With the legacy system, the total flux densities of their more compact sources matched those measured at OVRO or the UMRAO within a few percent. The VLBA data averaged about 4% lower than the single dish flux densities in a large collection of results they made available. When the transition was made to the RDBE using the PFB personality, the discrepancy went to closer to 25% and was rather sensitive to the bandpass normalization scheme used. Better results were obtained using normalization parameters that we really don't believe are correct. Meanwhile, we were processing science project BW102 which also used the PFB at 6cm (Wrobel, Walker, and Fu 2014). That project included OQ208, for which same-frequency data from the VLA were found from observations made 2 days later. OQ208 is slowly variable, so the VLA results were used to establish the expected total flux density of the VLBA image. A discrepancy similar to that found by MOJAVE was seen. In fact an offset of 31% was reported.

An effort was made to understand the nature of the problem using the BW102 data and data from several other existing observations. Comparisons of various bandpass calibration schemes were made. Opportunities were identified to compare the legacy system with the RDBE DDC personality and the PFB with the DDC. A new understanding of how bandpass calibration should be done was reached, but no effect was found that could explain an error as large as 25%. Therefore targeted test observations were made on 2014 August 14.

2: Bandpass normalization

The standard calibration sequence includes doing a bandpass calibration, although the importance of that step was probably not universally recognized. The importance is much increased with the use of filters with soft edges such as those in the RDBE[1] with the PFB firmware, especially when edge channels are not used in the processing. The difference between doing no bandpass calibration and doing a proper one, when using the inner 75% of channels for downstream processing is about 15%. The DDC firmware produces much sharper bandpasses as can be seen in Figure 1 where the two are compared. A sharp filter requires more resources on the FPGA. The PFB produces many soft baseband channels while the DDC produces fewer, but sharper baseband channels.

The recommended parameters for bandpass creation include a normalization step so that the bandpass does not alter the amplitude calibration. In the past, we recommended using the full band for the normalization (BPASSPRM(10)=4), but MOJAVE reported better results for the PFB using normalization over the inner 75% - the channel range used if the default for ICHANSEL was used. Having the calibration depend on the choice of channel range to use cannot be right. It would mean

1 The Roach Digital Backend (RDBE) is the new hardware on the VLBA that accepts two 512 MHz bandwidth Intermediate Frequency (IF) signals, filters to narrower baseband channels, constructs 4 level samples (2bit) and sends up to 2 Gbps to the Mark5C recording system. There are two RDBE units at each VLBA station allowing use of up to 4 IFs. The RDBE contains a field-programmable gate array (FPGA) for which 2 firmware personalities are available for common use. One is the PFB (Polyphase Filter Bank) which splits each incoming IF into 16 basebands of 32 MHz each. A total of 16 of the 32 basebands produced from the 2 IFs are recorded. The other personality is the DDC (Digital Down Converter) which provides up to 4 baseband channels with bandwidths between 1 MHz and 128 MHz each from each of the 2 RDBE units at each station.

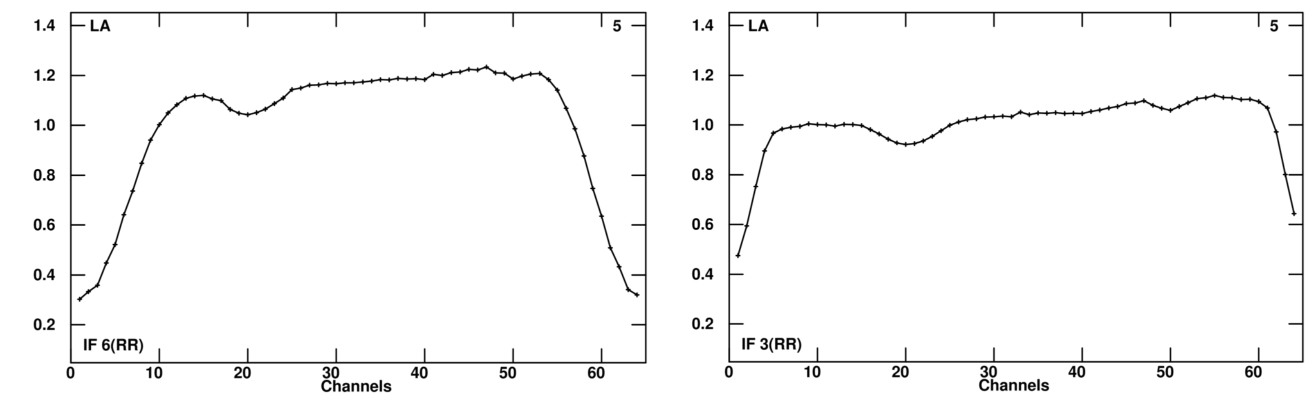

Figure 1: Autocorrelation bandpass shapes for the RDBE PFB on the left and the DDC on the right. These are for the test observation TA035 and are for the same 32 MHz bandwidth near 5 GHz. The data were taken about half an hour apart. Note the sharper edges for the DDC personality. Also note the difference in the levels of the central channels. ACCOR has been run so the average over all channels for each is unity. The similarity of the dips near channels 20 and 50 suggest that those features are in the IF and are not a product of the RDBE filters.

that the flux density scale depends on the choice of how many edge channels to toss, which it should not. There must be another reason for the calibration offset.

In the process of looking at the normalization, it was realized that the AIPS bandpasses are voltage gains for each channel. All AIPS gains are voltage gains to avoid having to take the square root when they are used to scale baseline data. When the autocorrelations are calibrated, they are multiplied by the square of the bandpass value for each channel because the ``baseline'' is between the antenna and itself. The normalizations applied by BPASS to the bandpasses were over the voltage gains. When applied to the autocorrelations, the desired result is that a normalized bandpass will not change the average value, which, after ACCOR, is unity. But with the voltage normalization, it did change the average value by about 2.7 percent. What is needed is a normalization that is based on the sum of powers (voltage squared). An option to normalize by power in BPASS was requested. Eric Greisen implemented it as options BPASSPRM(10) = 5 and 6. Option 5 can have restricted channels while 6 uses the full band regardless, which is the preferred option.

3: Results from existing observations

Various observations that were already in the archive were used to try to understand the calibration differences between the legacy and RDBE systems. This section outlines the important results obtained.

When the DiFX[2] correlator was commissioned, careful comparisons were done with the legacy correlator. It was determined that the amplitudes produced by both correlators were the same to better than a percent (Deller et al. 2007). Tests at NRAO showed amplitude differences of 0.79% ± 0.86% for autocorrelations and 0.39% ± 1.41% for cross correlations with 14,000 and 55,000 measurements respectively (Romney private communication). Thus, the change of the calibration offset was not the

2 DiFX is the software correlator that replaced the original hardware correlator on the VLBA. It enabled the use of wider bandwidths and has proven flexible in many ways.

Data for Bandpass	Firmware	Type	Normalization width	Average RCP	Average LCP
Auto Corr	DDC	No bandpass	NA	1.033	1.035
Auto Corr	DDC	Voltage	Inner 75%	1.032	1.034
Auto Corr	DDC	Power	Inner 75%	1.032	1.034
Auto Corr	DDC	Voltage	Full Band	0.996	0.996
Auto Corr	DDC	Power	Full Band	1.000	1.000
Auto Corr	PFB	No bandpass	NA	1.146	1.148
Auto Corr	PFB	Voltage	Inner 75%	1.147	1.149
Auto Corr	PFB	Power	Inner 75%	1.148	1.150
Auto Corr	PFB	Voltage	Full Band	0.973	0.972
Auto Corr	PFB	Power	Full Band	1.000	1.000
Cross Corr	DDC	Voltage	Inner 75%	1.047	1.052
Cross Corr	DDC	Power	Full Band	1.013	1.015
Cross Corr	PFB	Voltage	Inner 75%	1.163	1.169
Cross Corr	PFB	Power	Full Band	0.998	1.002
Cross Corr	PFB	Power no ACCOR	Full Band	0.976	0.980

Table 1: Effect on the average value of calibrated autocorrelations over the inner 75% of channels for various normalization options when forming the bandpass.

result of the change to DiFX.

Geodesy test observation TG009 was a comparison test of the legacy and RDBE DDC systems with both using 8 MHz bandwidths. A key result was that after ACCOR was used to scale the autocorrelations to 1.0 averaged across the band and FRING was used to flatten the phases with frequency to allow band averages, the cross correlation coefficients, averaged over all channels, are within 2% of each other with scatter between basebands somewhat larger than that. The legacy data have the slightly lower amplitudes but, with the restricted data examined, this is not significant. In any case, this result shows that the difference in calibrated amplitude scaling relative to flux densities measured with other instruments is a result of the calibration process, not the observing and correlation process.

Project TP028 was run on 2013 Jan. 28 on the VLBA. There were 20 segments which alternated between the PFB and DDC. The goal at the time was to stress the tools for changing firmware. But it also provided back-to-back data with the two systems on the same source (1357+769) very close in time. The DDC channels were picked so that they duplicated some of the PFB channels, allowing direct comparison. For the amplitude testing, TP028C (DDC) and TP028D (PFB) were loaded to AIPS. ACCOR and FRING were run on both data sets to remove any system offsets and to allow integration across the band. There were level changes during the first scan, so the reported results are from the second. Comparing the correlation coefficients integrated across the full band gave results that the DDC is about 1% higher than the PFB, which is too small to be responsible for calibration issues. As demonstrated above, the PFB has far softer edges to the baseband filters than does the DDC. If the average of the correlation coefficients without bandpass calibration is taken only over the inner 75% (BCHAN=16; ECHAN=112 when there are 128 channels), the result is that the PFB is about 11% higher than the DDC. This is an early indication that bandpass normalizations using the inner 75% are not going to be a good idea.

The TP028C and TP028D data sets were then used to compare various bandpass normalization options. Table 1 gives the average value of the inner 75% of the autocorrelation channels after ACCOR, FRING, and bandpass calibration with BPASS run with different normalization options. In this case, the result we want, as will be discussed more below, is to have the autocorrelation values in the channel range that will be retained to be equal to unity. For the PFB, typically only the inner 75% is retained because of the filter roll off. For the DDC, it is closer to 89%. From the table, it is clear that the most effective way to get a unity result is to form the bandpass from the autocorrelations using the power normalization over the full band. That normalization will assure that all calibrated autocorrelation channels end up with the same value – the average – and ACCOR forced the average to unity. All options involving normalization over a restricted channel range when using the PFB give big errors – 15 to 17%. That is far bigger than our advertised calibration errors of 5 to 10%. Doing no bandpass calibration is also a very poor option for the PFB and not so good for the DDC. Note that the difference between power and voltage normalization is much larger for the PFB than the DDC. This is a result of the soft filter edges in the PFB that are shown in Figure 1.

From Table 1, it looks as if we should recommend forming the bandpass from the autocorrelation data using power normalization over the full band. But, for reasons not related to establishing the flux density scale, one would like to have the bandpass include phases. You cannot get phases from the autocorrelations. If you calibrate with the cross correlations, you get phases, but also end up with a small offset from unity amplitude in the calibrated autocorrelations. That offset in the table is of order 1%. In other data, values of 2% were seen. To deal with this issue, a new AIPS task was written, called ACSCL, that is a close cousin of ACCOR, but that pays attention to the usual suite of data selection and calibration instructions. If ACSCL is run applying the ACCOR, FRING, and bandpass calibrations and selecting the inner channels that are going to be used for further processing, it will provide an SN table that, when interpolated to the CL table, will cause calibrated autocorrelations to be unity in the desired range. Note that, if ACSCL is used with solutions over the full time range of the data, it may not strictly be necessary to run ACCOR.

The difference between the scale determined with the cross and auto correlation derived bandpasses appears to be the result of aliasing that occurs, especially with the PFB, in the edge channels. The soft digital filters allow some signal from other frequencies to add to the signal near the filter edges. That added signal shows up in the autocorrelations and the system temperature values, but does not correlate. Bandpasses derived from the cross correlation data are low relative to the autocorrelations in the edge channels. Autocorrelation spectra calibrated with a cross correlation bandpass have large ``horns" in the edge channels. Those channels do not affect the final data used for imaging because they are tossed anyway because of the filter roll-off. When there is RFI, there may be other differences that users should be wary about, but those were not explored in this study.

4: The test observations

We were unable to understand the big calibration errors using existing data, despite the modified calibration methods that were implemented. It became clear that test data designed to check the amplitude calibration carefully and to compare the PFB, DDC, and Legacy systems was needed. A key aspect of the test runs was to obtain simultaneous, independent flux density measurements. Test project TA035 was designed to meet these needs. The VLA was included mainly to provide flux densities using internal data on the VLBI sources and the primary calibrator 3C286.

TA035 was actually four separate, back-to-back projects, all at close to 5000 MHz. TA035A used the PFB which always has 16 baseband channels, each of 32 MHz bandwidth. TA035B used the DDC with 8 baseband channels that matched 8 of the PFB channels in frequency and bandwidth. TA035C used the DDC and, at the two stations that still have them, the Mark5A legacy recording systems, all at 16 MHz bandwidth. TA035D was the same setup as TA035A. The phased VLA was used in the first two projects, using the 16 and 8 channel modes to match the VLBA. But more importantly, the VLA internal data were used to measure the flux densities of the VLBA target sources. The flux density scale on the VLA was set by 3C286. The sources observed were OQ208, J1224+2122, J1130+3815, and J1048+7143. Most of the comparison results were based on the last 2 sources. OQ208 is a double on the VLBI scale so obtaining an accurate total VLBI flux density was a bit tricky.   J1224+2122 had too much structure, even on the VLA for accurate comparisons.

There were various failures, so these observations did not give their full potential. The VLA had a bad module for TA035A, so we don't have VLBI data to the VLA for that observation. The VLA had poor weather during TA035B, so the flux density scale information comes exclusively from TA035A. The Mark5A (legacy) systems failed, so we do not have a fresh comparison with the RDBE. Otherwise, we did get good VLA flux density measurements from TA035A and we were able to compare data from the PFB and DDC personalities on the VLBA using TA035A/D and TA035B.

4.1: Flux density scale.

The VLA data were reduced using the CASA pipeline and using AIPS. There was a discrepancy of of about 2.5%.   This was traced mainly to a difference in the fundamental flux density scale used by the two processing paths. AIPS was using the Perley and Butler (2013) scale which is the current best practice. The CASA pipeline was using the 2010 scale, which is different by about 2.2%.   Fairly quickly after this issue was pointed out, the CASA pipeline was updated to use the 2013 scale. Correcting for that difference, the CASA and AIPS processing gave flux densities for the sources that differed by less than 0.4% on 3 of the 4 sources. The odd one out was J1224+2122 which seems to have other issues that make the results dependent on the details of the processing. It is not used in the current study.

The issues with the flux density scale on the VLA raised the question of what flux density scale is in use on the VLBA. The program that sets the flux density scale is PTANAL which is used to reduce pointing data to obtain the gains (in K/Jy) that are used, along with system temperatures, to calibrate flux densities. PTANAL uses polynomial coefficients to determine the flux density of some continuum sources. The coefficients were from Baars et. al. (1977) with modifications from 1990. The results for 3C286 are between the 2010 and 2013 scales. For high frequencies, DR21 and planets are

Data row	Expt.	Firmware	Bandpass / ACSCL	Channels (of 64)	J1130+3815 (Corr. coeff)	J1048+7143 (Corr. coeff)
1	TA035A	PFB	No / No	All	0.005739	0.007751
2	TA035B	DDC	No / No	All	0.005897	0.007905
	Ratio of data 1 over 2	PFB/DDC			0.973	0.981
3	TA035A	PFB	Yes / Yes	9 - 55	0.005864	0.007927
4	TA035B	DDC	Yes / Yes	4 – 60	0.005949	0.007973
	Ratio of data 3 over 4	PFB/DDC			0.986	0.994
	Ratio of data 3 over 1	PFB			1.022	1.023
	Ratio of data 4 over 2	DDC			1.009	1.009
5	TA035A	PFB	Yes / No	9-55	0.005733	0.007749
	Ratio of data 5 over 1	PFB			0.999	1.000

Table 2: Comparison of source correlation coefficients using the RDBE PFB and DDC with and without bandpass correction. The bandpasses are derived from the cross correlations and the final tweak to make the autocorrelations unity in the desired channel range (ACSCL) has been applied except in the final row. The final data row (no ACSCL) shows that the bandpass normalization has caused no change in average amplitude, as advertised. Thus the two ratios above for the same firmware with and without bandpass calibration are the factors applied by ACSCL. Note that the agreement of calibrated flux densities between the PFB and DDC in a later table is better than the correlation coefficient agreement. The Tsys values are slightly different.

used. The planet flux density calculations are made in code provided by Bryan Butler in 1996. Corrections were made for resolution for some sources, especially 3C274 and the planets. Clearly PTANAL is due for an update and one was made to the Baars sources to bring them to the 2013 scale. DR21 and the planets still need to be updated, but that should be done by the end of 2014. All of the sources used at 6cm were updated to the 2013 scale before the RDBE DDC based gains mentioned below were derived. That actually accounts for the bulk of the gain differences between the legacy system and RDBE DDC based gains.

4.2: Correlation coefficients.

Some comparisons were made using the correlation coefficients from the DDC and PFB, with and without application of bandpass calibration. That data had digital offsets removed with ACCOR and had the phase bandpass flattened using FRING. A bandpass was derived with BPASS using the cross correlation data and power normalization. ACSCL was use for a final tweak to the scaling to get unity autocorrelations in the channel range of interest, although one test was done without ACSCL to confirm that the bandpass normalization was working as advertised. Results of these tests are in Table 2. The scaling to flux density was not done for these tests so correlation coefficients are being compared. Task JMFIT was used to fit 2 component models to the data to obtain total amplitudes for comparison. With correlation coefficients, differences in system temperatures and differences in flagging between the data sets, of which there were some, can lead to differences in average correlation coefficients obtained by this method. On J1130+3815 and J1048+7143, amplitudes using full band integration with no bandpass correction match to less than 3%. The amplitudes, after bandpass calibration and the ACSCL tweaks, and leaving out the band edge channels, match to better than 1.5%. Comparing the data with and without bandpass calibration on TA035A without ACSCL shows essentially no changes, showing that the bandpass normalization on the bandpass derived from cross correlations worked as advertised. The final ratios for each firmware personality, with and without bandpass calibration, essentially shows the correction applied by ACSCL, which is basically the normalization difference between the auto and cross correlations. As expected, it is larger for the PFB with the soft filters than for the DDC. Note that, as will be shown below, after the full flux density calibration, the differences between the PFB and DDC drop to below about 0.4%, so slight differences in Tsys and probably improved amplitude fitting brought them into even closer alignment.

4.3: System temperatures.

The system temperature is an integral part of a priori calibration. The system temperatures, measured in the RDBE, for TA035A (PFB) and TA035B (DDC) were compared for the matching channels from the PFB and DDC. The differences averaged over channels and antennas to about -0.8% in RCP and -1.5% in LCP. Peak differences were about 4.6% with and RMS scatter of about 2% . Thus there were small differences in the PFB and DDC Tsys values. Judging by the improved agreement between calibrated flux densities (below) compared to the correlation coefficients, these small differences may be real.

There are significant differences between the system temperatures measured with the RDBE and the legacy system. Brisken (2011) reported at 6.5% difference between system temperatures measured with DiFX and the legacy system. The RDBE-measured system temperatures are closer to the DiFX-measured values, but I don't have a direct comparison. For both TA035 and BW102, the differences between the RDBE and legacy values averaged to about 8.2% with significant scatter between antennas and IFs. For pointing data taken between June and September 2014 and analyzed based both on legacy and DDC measurements (see below), a difference of 5.4% was found. Brisken (2011) argues that the RDBE values should be more accurate as they are based on simple digital statistics and they produce the lower Tsys values. Lower Tsys values require a higher difference between the power measured in the cal-on and cal-off states. Most of the ways one might imagine making an error in the measurements would reduce the cal-on/cal-off difference. These Tsys differences will directly affect the flux density calibration and could be involved in an explanation of differences between VLA and VLBA flux densities. We don't yet have a way to tell which of the RDBE or legacy system is correct.

Another issue related to system temperature is apparent in the calibrated data. While the correlation coefficients for the different baseband channels (AIPS IF/polarization) are very close, after calibration they is significantly more spread in flux density, at least for some stations. Examination of the tables of the cal temperatures (Tcal) used to convert difference power to system temperature shows point-to-point fluctuations between points measured on 50 MHz intervals that are probably too large. That will be reflected in Tsys variations with frequency that are too large, which in turn propagates to the flux density calibration. In the long run, the Tcal tables should be rebuilt. Wide band measurements suggest that the gains don't vary much across individual receiver bands, so, to simplify bookkeeping and maintenance, I would advocate picking a single gain and determining the Tsys, and hence, Tcal,

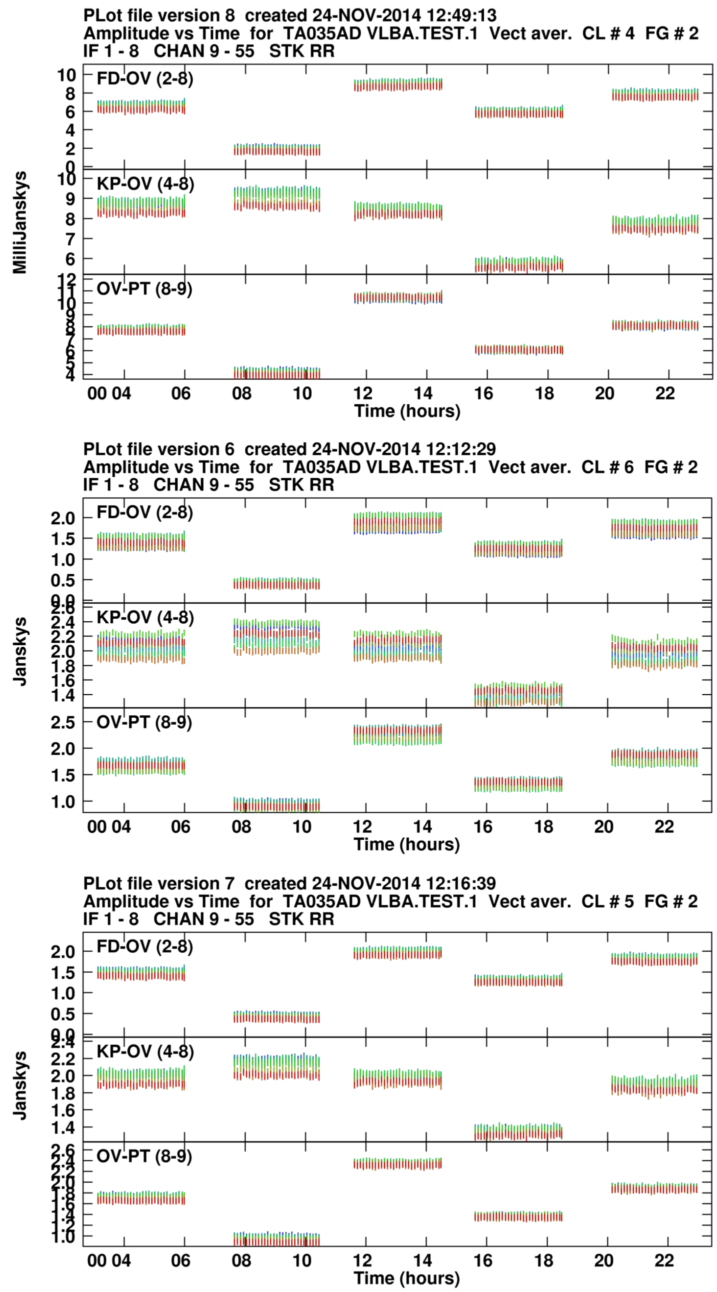

Figure 2: UV data from 3 baselines from TA035A (RDBE PFB). The top plot is bandpass calibrated correlation coefficients scaled with ACCOR and ACSCL. The middle and bottom plots are fully calibrated data after APCAL. For the middle plot, no attempt to smooth the Tsys values over IF was made. For the bottom plot, the Tsys values were shifted on average to match the channel at the frequency where the gains are determined. Note the spread of amplitudes in the middle plot, especially on KP-OV.

that is needed as a function of frequency to give the measured SEFD using that gain.

In the shorter run, a capability has been added to APCAL that helps deal with unwarranted fluctuations in the Tcal (and hence Tsys) values. APCAL can determine the average ratio over time of the Tsys for each baseband channel to the average over baseband channels, or to a chosen baseband channel, and then scale all the data for the baseband channel by the that ratio. That causes all baseband channels to have, on average, the same Tsys which is probably more realistic in the absence of RFI. The reason for the chosen baseband channel option is so that one can pick the baseband channel that matches the frequency used for the gain determinations – probably the preferred option so that the Tcal cancels out in the ratio of Tsys to gain. If there is RFI, this scaling option needs to be used with care. This capability is demonstrated in Figure 2. The top panel shows the calibrated correlation coefficients (ACCOR, BPASS, ACSCL). There is a bit of spread on KP-OV, but otherwise, not a lot of variation between baseband channels (colors). The middle panel shows the result after calibration with APCAL. Now the scatter between baseband channels is significantly higher because of the scatter of Tsys values caused by the scatter of the Tcal values. The bottom panel shows the values calibrated with APCAL with the option turned on to cause the Tsys values of all baseband channels to match, on average, the value of the channel that best matches in frequency where the gains are measured. Now the scatter is comparable to the scatter in the correlation coefficients, showing the value of this correction and the need to improve the Tcal values.

4.4: Updated gains.

The gains that are distributed in *vlba_gains.key* have so far been based on measurements made with the legacy baseband converters. Those are the gains provided with the archive data that make it into the AIPS GC tables. The system temperature values provided with correlated data are measured with the RDBE. Since there is observed to be an offset between the RDBE and legacy measured Tsys values, it would not be surprising to have an offset in the gains too. Also there is the issue of the flux density scale.

The ability now exists to do a pointing analysis based on power measurements made with the RDBE with the DDC personality. Walter Brisken has provided a version of the FIT program that runs on RDBE data. FIT is the program that converts power data made during pointing patterns to pointing offsets, antenna temperatures, system temperatures and all the other auxiliary data required for pointing analysis. After updating PTANAL's flux density scale for lower frequencies, including 5 GHz, pointing data from June through September 2014 were combined and analyzed for three cases. One was based on the legacy baseband converter data and the old flux density scale. Another used the RDBE/DDC power data but the old flux density scale. The third used the RDBE/DDC power data and the new flux density scale. As mentioned above, the system temperatures, averaged across stations and polarizations, were 5.4% lower for the RDBE than for the legacy system. The flux density scale did not enter the Tsys calculation so the two RDBE cases were the same. This legacy - RDBE Tsys difference is smaller than seen in Brisken (2011) or what was seen in TA035 and BW102. The significance of the difference is not clear. The gains, however, from the legacy and RDBE system, both determined using the old flux density scale, match to within 1% with the gains measured with the RDBE being slightly smaller. The scatter between stations and polarizations is more like 4% so the gains are effectively the same. When the 2013 flux density scale is used with the RDBE based data, the gain offset between the old flux density scale and the new one is 3.5% with the new scale gains being higher.

In hindsight, it is not too surprising that the gains did not change much between the legacy and RDBE data using the same flux density scale. The analysis system obtains the gain in degrees K per Jansky by comparing the power change going onto source with the power change when the cal is fired. The sources used are not too different in contributed power from the cal so the changes are over a similar fractional range of powers. Even if one or both of the systems is non-linear or has a zero offset, the relative power change for the source and cal will be similar. The same cannot be said for the ratio of the power change when the cal is fired to the total power, which are the data on which the system temperatures are based.

Changing the gains in an AIPS data set is a bit painful. The way it was done for the tests reported here was to write out the GC table with TBOUT, edit the result with a text editor or a sed script, then read it back in with TBIN. Later it was confirmed that the job can be done using VLOG to extract gains from *vlba_gains.key* and ANTAB to read the results into a new GC table. Those are AIPS tasks that were used to read the calibration tables before calibration transfer was implemented.

4.5: Opacity and Spillover

Another area where issues were found with current practice was in dealing with opacity. The gains in *vlba_gains.key*, and hence the GC tables, are "opacity corrected". That means that they are meant to apply to the flux density actually received at the antenna after any atmospheric absorption. An uncorrected gain would include the effect of the absorption and apply to the flux density of the source in free space. That has the disadvantage that the opacity is time and elevation dependent and so cannot be described as a constant. Over most of the bands of the VLBA, the opacity is roughly 1%, which shows up as a zenith atmospheric temperature of about 3 degrees. The derived opacity corrected gains are thus 1% (or often a bit more) higher than the uncorrected gains. This is because the gain is the ratio of antenna temperature to flux density. Antenna temperature is the measured result and doesn't change, but the flux density of the calibrators corrected to what is received at the antenna is lower than the free-space value.

In AIPS, it is common to not do the opacity corrections at 2cm and longer wavelengths. But that does invite an error. That error is not large in good weather, but it is there. AIPS task APCAL can derive the opacity using a Sec(Z) fit to the system temperatures measured during the run. From that fit, a "receiver" temperature is derived (that also includes non-receiver terms – any terms that are constant). Any excess over the receiver temperature, on a scan by scan basis, is assumed to come from opacity. The value of opacity is derived by comparing the added system temperature to a temperature derived from the weather station data, adjusted downward by some amount to roughly account for the fact that most of the absorption occurs well above the surface. Note that RFI can confuse this process.

There is an additional elevation dependent term to the system temperature which is accounted for in APCAL. That is spillover. The spillover tends to be a bit high at the zenith, when lines-of-sight that pass the upper edge of the dish come onto the ground. Those paths quickly move onto cold sky as the antenna tilts. At low elevations, there is more serious spillover as the lines-of-sight from the feeds that just miss the subreflector come onto the ground. If there are any odd optics, such as the dichroic plate over S band on the VLBA, or the lens over the old VLA L band feed, there can be additional spillover. It is difficult to measure the spillover because it it hard to separate from the atmospheric effect. For the VLBA, it was done early in the VLBA history by looking at the deviations from a clean Sec(Z) law of the Tsys with elevation. But the numbers are just estimates. APCAL does apply the spillover corrections. Unfortunately, it was found while working on the projects reported here that APCAL used the spillover function derived for the 7mm receiver for all receivers. That was not appropriate and, in fact, often led to a derivation of zero opacity for the lower frequency bands. PTANAL uses band dependent spillover curves. Those curves were provided to Eric Greisen who implemented them in APCAL.

4.6: Results for TA035

Table 3 gives the results from the TA035 observations for the two sources for which resolution effects were not a problem plus OQ208 for which zero spacing flux densities could be eyeballed from plots. The sources were a bit resolved, as usual for VLBI. The flux densities for J1130+3815 and J1048+7143 were extracted from the calibrated UV data using an amplitude-only fit for two components (JMFIT). This measurement method was compared to other schemes and little difference was found. OQ208 is a double and shows rather stronger resolution effects. The VLBA flux densities in the table are eyeball estimates of the zero spacing values from a plot of calibrated, but not self-

Source	J1130+3815	J1048+7143	OQ208
VLA flux density (Jy)	1.407	1.952	2.14
VLBA RDBE PFB flux density (Jy) TA035A and TA035D combined	1.335	1.880	2.14
VLBA RDBE PFB flux density (Jy) TA035A only	1.345	1.884	
VLBA RDBE DDC flux density(Jy) TA035B	1.344	1.876	2.14
Ratios:			
Ratio VLA / VLBA PFB (TA035A+D)	1.054	1.038	1.00
Ratio VLA / VLBA PFB (TA035A)	1.046	1.036	
Ratio VLA / VLBA DDC	1.047	1.040	1.00
Ratio VLBA PFB/DDC (TA035A&D)	0.993	1.002	1.00
Ratio VLBA PFB/DDC (TA035A only)	1.001	1.004	

Table 3: Results from test observation TA035 comparing VLA flux densities with VLBA values with the RDBE with the PFB and DDC personalities. Bandwidths were 32 MHz in all cases. The VLBA results are from 3 matching baseband polarization pairs (AIPS IFs). The VLA results are for the same center frequency as the VLBA data. The results for J1130+3815 and J1048+7143 are from 2 component fits to the UV data amplitudes. The results for the more resolved source, OQ208, are from eyeball estimate of the zero spacing flux density based on plots of the flux density vs UV distance (the perfect match is surely fortuitous). Note the extremely good agreement between the PFB and DDC data.

calibrated, amplitude vs UV distance (UVPLT). The key result is that the VLBA shows flux densities 4 to 5 % lower than the VLA for the two slightly resolved sources and is closely matched to the VLA in the less certain measurements of the resolved source OQ208. This is far closer than the offsets that triggered this study. The reasons for the remaining offsets in TA035 are not yet known. Possibilities include some resolved structure for these sources, Tsys measurement problems, a coherence issue, or some digital loss factor ("b factor") that is not being accounted for. The other result apparent here is that the RDBE and DDC agree to less than half a percent when only considering TA035A and B, and about 0.7% for the biggest difference using both TA035A and D for the PFB data. Thus we have very consistent calibration between the PFB and DDC.

4.7: Results for BW102.

The result for BW012 are not quite as encouraging, although certainly better than the 31 % offset reported in Wrobel, Walker, and Fu (2014). Table 4 gives the flux densities measured after dealing with all the modifications to the calibration procedures discussed in this memo. The VLA flux density measured on July 1, 2013 is the one used for the comparison in Wrobel, Walker, and Fu (2014). Two additional VLA measurements were found from before BW102. Table 4 gives comparisons with all 3 VLA results. We were a bit surprised at the magnitude of the OQ208 flux density variations seen in the VLA data, but don't have an explanation.

BW102 OQ208	Flux density	VLA/VLBA (A&P self-cal)	VLA/VLBA (P self-cal)	Elevation (deg) OQ208	Elevation (deg) 3C286
VLBA (A&P self-cal)	1.958				
VLBA (P self-cal)	2.017				
VLA June 22, 2013	2.220	1.134	1.101	22	50
VLA June 23, 2013	2.180	1.114	1.081	42.5	28.5
VLA July 1, 2013	2.287	1.168	1.134	70	73

Table 4: VLBA flux densities for OQ208 in observation BW102 on June 29, 2013. The VLA flux densities are from archival data nearby in time. The calibration strategy recommended in this memo has been followed for the VLBA. The VLA flux densities are in the Perley Butler (2013) scale.

The difference between the VLBA amplitude and phase (A&P) self-calibrated result and the phase only (P) self-calibration result is related to a common tendency for amplitude self-calibration to draw down the short baselines and raise the long baselines. This is mainly an issue with a relatively small number of antennas and, especially, with the antennas supporting the long baselines not participating in short baselines. It is an issue for the VLBA, but far less so for the VLA. Note that the eyeball scheme for obtaining the zero spacing flux density from an amplitude vs UV distance plot, such as was used on this source for TA035, gave 2.01 Jy, insignificantly different from the phase-only self-calibration result. For a source with structure such as OQ208, the phase-only (or eyeball) result is probably a better estimate of the total flux density than the total in an image after amplitude self-calibration, so the offsets between 8.1% and 13.4% are considered here to be the indicator of the error in the VLBA calibration.

These offsets are significantly higher than what was obtained in TA035. The reason is not understood. The difference between the system temperatures in BW102 between the legacy and RDBE systems is about 8.1% on average. If the legacy system temperatures are the correct ones, that would bring the VLA-VLBA difference into a tolerable range. But I am not aware of any change in the Tsys measurement scheme between BW102 and TA035. Indeed, the difference between the legacy and RDBE Tsys values for TA035 is about the same as for BW102. But the calibration of TA035 seems to have worked reasonably well. Perhaps there was some reason for reduced coherence in the older data (BW102 was about a year before TA035 and the RDBE system is under continuing development), but I am not aware of anything that was changed that would cause such an effect.

At this point, it would be useful to have MOJAVE recalibrate some of their data using the recommended style to see if they still have issues Presumably the offset between their VLBA images and the OVRO flux densities will get smaller, but they may still be significant. It would also be useful to gather statistics on the VLA/VLBA difference.

5: Recommended amplitude calibration procedure

The new recommended procedure for amplitude calibration is given here. The first steps remove the instrumental effects and prepare the correlation coefficients for multiplication by the SEFD to obtain correlated flux densities. The goal of these first steps is to derive a bandpass and data scaling factors that, together, give a value of unity to the calibrated autocorrelations across the channel range of interest and over time. Then the multiplication by SEFD, which involves the system temperature and gain, is done. As part of that step, atmospheric opacity corrections should be made.

Step 1: Scale the data to correct for digital offsets in the sampling and correlation. This is done by deriving scaling factors as a function of time that cause the average value, over all channels, of the autocorrelations to be unity. In AIPS, use task ACCOR. The solution interval should be short compared to any variations and any interpolations should not be done across sources. Among the effects being corrected are the effect of variations in the total power reaching the samplers which can cause changes in the relative occupation of the various sampler levels given the granular nature of the on-line level setting. Thus events like source changes are likely to be times when there are abrupt changes.

Step 2: Prepare to derive a bandpass by removing the residual delay, which shows up as a phase slope, in the data which will be used to make the bandpass. This can be done as part of applying the pulse calibration data (AIPS task PCCOR) or with a fringe fit (AIPS task FRING). One does this using tools that will describe the phase slope as a delay rather than just including the slope in the bandpass. Any correlator on-line averaging results in some loss of amplitude when there is a phase slope. In AIPS, corrections are made for this effect if the slope is described as a delay but not if it is just a slope in the bandpass. This amplitude loss is non-closing which means that it is baseline dependent and cannot be removed with self-calibration. Hence, if not corrected, it can limit dynamic range.

Note that any operations that adjust phases on any basis other than applying a single change across the full time of the observation should be done after a priori phase corrections such as ionospheric corrections or EOP updates. Otherwise a fringe fit or self-calibration, for example, might have already removed the effects to be corrected and you end up making the correction twice.

Step 3: Make a bandpass (AIPS task BPASS). After the autocorrelation scaling above (ACCOR), using the autocorrelations as the basis of the bandpass would automatically give the desired final result for the scaling. However one also wants the phase bandpass, and it is impractical with current software to determine that after doing the amplitudes with the autocorrelations. So the bandpass should be made using the cross correlations on a strong source. The normalization should be over the full bandwidth. Watch out for the default channel selection being the inner 75%. You really don't want that as it will lead to errors upwards of 15% with the soft edged filters of the RDBE PFB personality. The normalization should also be based on power as discussed in Section 2 above. The normalization of the cross correlation bandpass is not quite the same as the autocorrelation bandpass because of aliased signals near the band edges in the autocorrelations. Those signals do not correlate, so the cross correlation bandpass does not flatten the power in that region of the autocorrelations. There might be similar issues with RFI, but that was not addressed in this study. The non-correlating edge channels are outside the channel range usually of interest so are not harmful. But they do lead to the need for the final amplitude adjustment of the next step.

Step 4: After the steps above, the calibrated autocorrelation amplitudes in the region of interest will be slightly offset from unity. Values are typically near 0.98 with the RDBE PFB personality and 0.993 with the RDBE DDC personality. Correcting this is a small adjustment, but an easy scheme has been provided to accomplish the correction. AIPS task ACSCL is an enhanced version of ACCOR that has the data selection and calibration options available in other programs. One can select the channels that will be used in imaging or other post processing and have the average value of the autocorrelations determined in that range. The correction factor to calibrate the autocorrelations over the desired range are then derived. If doing Step 4, it might be possible to skip Step 1 (ACCOR), but that has not been explored.

After Step 4, the data are properly scaled for the final conversion from correlation coefficients to flux density. This study did not focus on what happens in the presence of RFI, so, if you have significant RFI in your data you should keep a close eye on the process. Proper calibration would depend on whether the Tsys values properly reflected the additional power from the RFI. Unfortunately, the Tsys values will usually be corrupt when there is significant RFI so calibration gets trickier.

Step 5: The cross correlation coefficients after the above calibrations are converted to flux density by multiplying by $\sqrt{SEFD_1 \times SEFD_2} = \sqrt{(T_{s1} \times T_{s2}) / (g_1 \times g_2)}$ where the *SEFD* is the system temperature (T_{si}) divided by the gain (g_i). In AIPS, this is done by APCAL. As discussed earlier in this memo, the opacity corrections offered by APCAL should be used for the best calibration. Also, the effects of Tcal variations can be corrected by invoking the option to scale all the Tsys to match, on average, across the IFs. If there is significant RFI, or a baseband goes outside some filter in the system, this needs to be done with caution as the RFI may introduce legitimate Tsys variations. The gains used should be those determined using the RDBE DDC data and using the Perley & Butler (2013) flux density scale. As of this writing, the gains distributed with data still use the old flux density scale, but that should change shortly.

This completes the a priori flux density calibration. The alignment of the calibrations of different stations can be improved by imaging a calibrator and using the model to drive adjusted gains using self-calibration. Those gains can be applied to all sources. While doing any amplitude self-calibrations, it is important to think about the option to restrict the gain normalization to a subset of the antennas in the solution (ANTUSE in CALIB). Data from a station can be perfectly usable in self-calibration, but may have poor absolute calibration because of weather, pointing, or a new receiver for which well determined gains are not yet available. Such antennas should be included in the self-calibration, but should be excluded from the gain normalization – they should not be allowed to influence the overall flux density scale because they are not well calibrated.

Appendix A: List of issues found in this study

This is a summary of the problems found, and procedures modified as a result of this attempt to explain the large flux density offsets found by MOJAVE and BW102.

1. The CASA pipeline was still using the 2010 flux density scale. This has been fixed.

2. The VLBA gain determinations were using a 1990 flux density scale. This is in the process of being fixed, but the gains in active use are still on the old scale. The process of updating the AIPS gain table is a bit awkward and could, perhaps, be improved.

3. There is a significant difference (5 – 8 %) between the RDBE and legacy system temperatures. This is not yet understood.

4. The bandpass normalization should be done on power, which wasn't an option before this project. It was added to BPASS.

5. The bandpass normalization should be done over the full band. This was the common wisdom before this effort, but there were some who noticed that omitting the edge channels gave better calibration. That was presumably for other reasons.

6. The goal of bandpass calibration plus calibration for digital offsets should be to have the average value of the calibrated autocorrelation be unity.

7. To achieve the unity autocorrelation goal, and still use cross correlation data to determine the bandpass in order to include phases, the scale needs to be tweaked. New AIPS task ACSCL was provided for this.

8. Before ACSCL was available, I used CLCOR to scale the gains. I found that the gain curve option in CLCOR applied the polynomial curve as a voltage gain. The telescopes are almost certainly providing power gain curves. A power gain curve option (POGN) was added to CLCOR.

9. The current AIPS procedure VLBACALA runs ACCOR and APCAL back to back. With the suggested calibration steps, this is not appropriate as PCCOR or manual PCAL (FRING), bandpass calibration, and ACSCL should be run between the two tasks. Also, as per current recommendations, it is probably best to run the ionospheric corrections and EOP updates before the PCCOR/FRING step.

10. We need to encourage the use of opacity corrections since the gains we deliver are for that case.

11. The spillover corrections used for the derivation of opacities in APCAL were inappropriate at most bands (all but 7mm). That was corrected, and improved the results.

In the end, the calibration of the test project TA035 was successful. But the older projects still have offsets. The reason is not clear. It should be identified.

Appendix B: To do list

Here are a few issues that could/should be addressed.

1. The conversion of PTANAL to the Perley & Butler (2013) scale needs to be finished. This mainly means updating the derivation of flux densities for DR21 and for the planets.

2. DDC based gains in the 2013 scale need to be derived and made available to the users.

3. A less tedious scheme for updating GC tables is needed. It would be nice to just read a new *vlba_gains.key* file.

4. We still don't understand the large calibration errors in MOJAVE in BW102. They are smaller with the new calibration sequence than what was originally found, but are still considerably larger than the generally assumed accuracy of 5% or so.

5. The difference between the legacy and RDBE Tsys values should be understood.

6. The AIPS Cookbook and AIPS VLBA scripts should be updated to the new scheme.

7. Statistics should be gathered from other projects on the difference between VLA and VLBA flux densities to understand the small remaining differences. Perhaps, in the end, we should recommend using a small "b factor" (APARM(1) in APCAL) to correct a difference if it proves consistent, even if we don't fully understand it.

8. The Tcal tables need work. This is a future project, but has become more important with the wider bandwidth data sets. Meanwhile users should be encouraged to use the feature that evens up the Tsys values across AIPS IFs, preferably to match the IF that is closest to the frequency at which gains are determined.

Acknowledgments:

I would like to thank Drew Medlin for processing the VLA data for the 2013 projects through the CASA pipeline. I also thank Walter Brisken, Vivek Dhawan, Amy Mioduszewski, Jon Romney, and Joan Wrobel for useful discussions about the issues in this study, Emmanuel Momjian for help with processing the test observation in the CASA pipeline, and Eric Greisen for being responsive to my requests for modified capabilities in AIPS. Finally I would like to thank Matt Lister, Yuri Kovalev, and the MOJAVE project for calling the calibration problem to our attention. This work made use of the Swinburne University of Technology software correlator (DiFX), developed as part of the Australian Major National Research Facilities Programme and operated under licence.

References:

Baars, J. W. M., Genzel, R., Pauliny-Toth, I. I. K., 1977, A&A, 61, 99.

Brisken, W. VLBA Upgrade Memo #34. (2011), "System Temperature Determination with DiFX"

Deller, A. T., Tingay, S. J., Bailes, M., and West, C. 2007, P.A.S.P., 119, 853.

Perley, R. A. & Butler, B. J., 2013, Ap. J. Supp. 204, 19.

Wrobel, J. M., Walker, R. C., & Fu, H. 2014, Ap. J. 792, L8.