Technical Implementation Plan for the Jansky Very Large Array Sky Survey (VLASS)

Steven T. Myers, Claire J. Chandler, & Casey Law

November 4, 2014

Contents

1 Overview

This working document details the Technical Implementation Plan (TIP) for the VLA Sky Survey (VLASS).

The main Survey Science proposal "The Jansky-Very Large Array Sky Survey (VLASS)" contains the science justification and overall description of the survey. This supplementary document describes in more detail the techncial issues, project plan, and risks associated with the proposed VLASS.

Version Note: This draft contains slightly revised numbers for the needed time for the survey and a small number of other changes, compared to the Survey Science proposal draft submitted on Oct 15[1]. The numbers in this TIP supersede those in the draft, and will be incorporated into the next draft of the survey proposal.

2 Basic Assumptions

The following assumptions are used to calculate the parameters of the survey.

2.1 Point-Source Sensitivity Assumptions

The point-source sensitivity of the Jansky VLA for mosaicking is computed using the procedure given in the Guide to VLA Observing: Mosaicking[2]. For continuum (Stokes I) at S-band (2–4 GHz) we assume a Survey Speed (SS) of

$$SS = 16.55 \left(\frac{\sigma_I}{100\,\mu\mathrm{Jy/beam}} \right)^2 \ \mathrm{deg}^2\,\mathrm{hr}^{-1} \tag{1}$$

of on-sky integration time for an assumed image rms of σ_I (μJy/beam). This assumes 1500 MHz of useable bandwidth (after RFI excision) and an image averaged over the band using multi-frequency synthesis. The integration time needed to survey a given area to a depth is given by dividing that area by the survey speed.

2.1.1 Useable bandwidth

A key parameter that factors into the sensitivity is the amount of useable bandwidth from 2–4 GHz. Standard JVLA guidelines and the VLA Exposure Calculator Tool[3] recommend a value of 1500 MHz, and therefore this is what we used in Equation 1, and was employed to calcualate the required observing time to reach a given depth. However, the exact amount of RFI affected frequency space is not yet carefully quantified, and also depends upon direction (antenna pointing direction in Azimuth and Elevation in many cases).

As a case in point, the Stripe-82 surveys referenced throughout this TIP (12A-371 PI: Kulkarni, and 13B-370, PI: Hallinan) employed fairly brutal excision, throwing away whole spectral windows. The effective bandwidth used for the imaging of these datasets was around 1350 MHz. If the VLASS employed this level of bandwidth reduction, then we would require an 11% increase in required observing time to reach the specified depths of the survey, or would incur a 5.4% penalty in acheived rms imaging sensitivity (e.g. 105μJy/beam for ALL-SKY instead of 100μJy/beam).

Risk: Medium. We consider this to be medium risk for the VLASS. This is one of the key issues to be resolved in the Test and Development Program (§ 10).

---

[1] `https://safe.nrao.edu/wiki/pub/JVLA/VLASS/VLASS_subm1.pdf`

[2] `https://science.nrao.edu/facilities/vla/docs/manuals/obsguide/modes/mosaicking`

[3] `https://obs.vla.nrao.edu/ect`

2.1.2 Sensitivity Loss at Low Elevation

The VLA exhibits sensitivity loss at low elevations, in the S-band primarily due to an increase in system temperature from spillover from the over-illuminated secondary. In the calculations used to estimate the required integration time, we have used measured values for this noise increase provided by Rick Perley (private communication).

A more detailed explanation of the sensitivity and observing time calculations are available in a online iPython Notebook.[4]

2.2 Angular Resolution

The Observational Status Summary (OSS)[5] for the VLA lists the angular resolution (FWHM) for S-band (at 3GHz) as 2.1″ in B-configuration and 0.65″ in A-configuration. These are meant to be indicative for long-tracks at high declination using *uniform* weighting (equivalent to a fully filled uv-plane out to the maximum baseline). The current advice given to users in the OSS is "The listed resolutions are appropriate for sources with declinations between -15 and 75 degrees...The approximate resolution for a naturally weighted map is about 1.5 times the numbers listed for θ_{HPBW}. The values for snapshots are about 1.3 times the listed values." These values are derived from the 1995 PhD thesis of Dan Briggs[6] based on classic VLA tests using the narrow continuum bandwidths then available. More careful examination of the Briggs thesis suggests that the resolution for naturally-weighted snapshots is 1.4 times the numbers listed for uniform long-track observations in the OSS.

The above standard numbers for uniform weighting, coupled with the Briggs factor of 1.4 for naturally weighted snapshots, imply that we should obtain a naturally weighted resolution of 2.9″ in B-configuration and 0.9″ in A-configuration. These are expected to be valid for regions with $-15^\circ < \delta < 75^\circ$, fields outside this range will have degraded resolution (unless observed in a hybrid configuration).

It is not clear how applicable the Briggs thesis numbers are to the wideband JVLA obseravtions proposed here, and so this issue will be further investigated as part of the Test & Development Program (§ 10.1.2). In particular, the Briggs thesis found that "super-uniform" weighting of snapshots would yield a rms sensitivity 1.11 times the naturally weighted values, with an increase of the beam size of only 1.2 times that for uniform weighting. If this holds for VLASS continuum imaging, this would yield resolution of 2.5″ in B-configuration and 0.8″ in A-configuration at a rms image sensitivity level 1.11 times the natural weighting values quoted for the survey components in § 3. It is however the baseline plan to produce naturally weighted images in the Basic Data Products, differently weighted images would fall under Enhanced Data Products (see below).

2.2.1 Case Study: Stripe-82 and COSMOS at $\delta = 0^\circ$

In the 13B-370 Stripe-82 B-configuration data[7], we found (at Declination $\delta = 0^\circ$) snapshot synthesized beam sizes (FWHM) of around $3.4'' \times 2.3''$ for natural weighting, with a geometric mean beam size of 2.8″. This is encouragingly close to the expected values based on the Briggs study.

---

[4] `http://goo.gl/jHEPdQ`

[5] `https://science.nrao.edu/facilities/vla/docs/manuals/oss/performance/resolution`

[6] `http://www.aoc.nrao.edu/dissertations/dbriggs/`

[7] SB `13B-370.sb28581653.eb28626177.56669.781848645835`

2.3 Overhead Assumptions

In the estimates for total observing time, the assumption is made that there is a overhead for slewing, setup, and calibration applies to the components of the survey. For example, for general VLA observing, we recommend use of a 25% overhed, where integration times are to be multiplied by 1.25 in order to arrive at the ”clock time” needed to execute observations. In practice, the overhead will depend on exactly how the survey components are scheduled and how much calibration is required, and what calibration can be shared between blocks.

Below we present two case studies based on an actual schedule and on a simple model respectively. Based on these studies, we fully expect to be able to meet (or come in under) the global 25% overhead for (i) observing blocks of 6 hours or longer, and (ii) short blocks that can share calibration with other blocks observed around the same time, (iii) blocks of 3 hours or longer that are fixed-time scheduled. More detailed verification of overheads, including full calibration, are scheduled as part of the Test Plan.

2.3.1 Case Study 1: Stripe-82

For example, the S-band Stripe-82 observations of program 13B-370 (Hallinan et al.) were observed using dynamic scheduling and independent 3-hour blocks with 2.25 hours of on-target observing, and thus had a 33% overhead. This particular base schedule were self-contained, and included standard calibrators 3C48 (for flux density and polarization angle) and 3C84 (low-polarization leakage calibrator). The OTFM scans consisted of 15 stripes to cover the Declination range, with each stripe covering 36min (9 degrees) in RA. At the scan rate chosen, each stripe took 9.25 min to get 9 min on-source integration (3% scan overhead). Groups of two stripes were interspersed with calibrator scans (with a singlet stripe left over in the block). There were extra observations of calibrators from adjacent blocks included to help link the blocks. This was a fairly conservative strategy, but control of calibration errors was important for the program.

Starting with the actual SBs used for the observations (from 1, 2 or 3 blocks for the different lengths), a number of example SBs were constructed under a range of different assumptions. The overheads ranged from 24% to as low as 14% under different less conservative criteria. See Table 2.3.1 for results from the example schedule construction exercise.

Overheads for Example Scheduling Blocks

Duration (hours)	**Dwell (hours)**	**Overhead**	**Dynamic?**	**Notes**
3.0	2.25	33%	yes	(1)
2.75	2.25	22%	no	(2)
5.6	4.5	24%	yes	(3)
5.25	4.5	17%	no	(2)
7.7	6.75	14%	no	(2)

All blocks contained full calibration. Note that whether an example is ”dynamic” or not depends on whether this SB could be submitted over a signficant range of LST and thus includes extra padding for possibly longs slews or wraps. If it needed a narrow range of starting LST then it was deemed to be essentially a fixed time schedule.

Table Notes: (1) 13B-370 original schedule, multiple gain calibrators, calibration every 20min; (2) single gain calibrator, calibration every 30min; (3) multiple calibrators, calibration every 20min.

2.3.2 Case Study 2: A simplified model

We now present a simplified model of overheads in SBs, guided by the first case study above. In this model we assume:

- a fixed startup overhead (to slew to first source) of 10 min.
- a fixed time on per calibrator observation of 2 min.
- for each hour of on-source integration, a number N of 2 min. calibration scans (e.g. N=2 or N=3)
- a fixed flux density and polarization calibration (on a low polarization source) of 15 min.
- (if needed, an additional fixed polarization leakage calibration of 15 min with 3 scans at different PA with 5 min. per scan)

Thus in this model, there is a fixed overhead of 27 minutes, plus a possible additional overhead of 15 minutes for the polarization calibration, so the fixed per-block overhead is either 30 minutes or 45 minutes (rounded). There is also 4min./6min. per hour of on-target time.

These assumptions are meant to be appropriate for large areas over a range of Declinations, where slew times to calibrators may vary. Small areas (e.g. Deep fields) may or may not conform to this model.

Assuming our calibration every 20 min (N=3), and allowing for a single leakage scan on an unpolarized source (or use of the phase calibrators for PA coverage of a polarized source), then we get the following costs:

Overheads for Simplified Model A — Calibration every 20min, single leakage scan

On-source time (hours)	Duration (hours)	Overhead
3.0	3.75	25%
4.0	4.75	19%
6.0	7.05	18%
10.0	11.45	15%

The overhead for calibration every 30 min. (N=2) is about 3% lower in all cases.

If we add the additional 15 min. fixed overhead for 3 leakage scans, then we get:

Overheads for Simplified Model B — Calibration every 20min, 3 leakage scans

On-source time (hours)	Duration (hours)	Overhead
3.0	4.00	33%
4.0	5.00	25%
6.0	7.30	22%
10.0	11.70	17%

The overhead for calibration every 30 min. (N=2) is again about 3% lower for all cases.

2.3.3 Applying the Simplified Model to VLASS

We will use Simplified Model B as a conservative estimate of overheads for the VLASS.

We assume that:

Overheads for VLASS under Simplified Model B

Component	Overhead	Comments
ECDFS	25%	5-hr blocks, contained calibration
COSMOS,E-N1,GALACTIC	22%	6–8 hour blocks, contained calibration
ALL-SKY, WIDE	17%	long blocks, shared calibration

These numbers are adopted as guidance in the per-Tier times given below. Note that the overheads actually used for scheduled hours for the DEEP fields are close to these, but rounded to make the schedules of length rounded to nearest 15min.

It is likely possible to further reduce overheads, particularly for the DEEP fields which are observed repeatedly over multiple days per cycle, by sharing primary calibration (flux density, polarization angle, possibly leakage). This will require testing to determine the optimum strategy.

3 Survey Structure

The VLASS science case is comprehensive and motivates a multi-tiered survey consisting of three tiers at different depths: ALL-SKY (Tier 1) at 100 μJy rms, WIDE and GALACTIC (Tier 2) at 50 μJy rms, and DEEP (Tier 3) at 1.5 μJy rms. The observing time required for the three tiers is 1904 for ALL-SKY (exclusive of WIDE), 2824 hours for WIDE, 840 hours for GALACTIC (exclusive of ALL-SKY), and 3391 hours for DEEP. The total scheduled observing time in VLASS, including overhead, is 8959 hours for our primary, preferred, 5-year 4-cycle schedule.

3.1 Tier 1: ALL-SKY

The ALL-SKY tier covers the entire sky visible to the VLA north of Declination -40°, 33885 square degrees (for 82.14% of the sky), with 23885 square degrees exclusive of the WIDE area. The observations will be carried out using On-The-Fly Mosaicking (OTFM). The scan rate will be adujsted depending on declination (elevation) to provide uniform sensitivity, requiring an extra 9% net integration time over the entire area. To reduce risk, we propose as a baseline plan to observe this tier in a single pass.

The summary statistics of the ALL-SKY tier are:

- B-configuration (2.7$''$ resolution)
- total visible sky area of 33885 square degrees, $\delta > -40^\circ$
- combined continuum image rms (Stokes I) $\sigma_I \geq 100\,\mu$Jy
- effective integration time required is 2047.4 hours
- multiply by factor 1.09 for sensitivity loss at low elevation for low declinations (§ 2.1.2), requiring 2232 hours true integration to reach uniform sensitivity over entire area
- area of 23885 square degrees exclusive of region covered in Tier 2 WIDE
- true integration time in this area (excluding 1 epoch of WIDE, 604 hours) requires 1628 hours
- total of 1904.8 hours observing required with 17% overhead
- total Tier 1 (exclusive) scheduled time of 1904 hours
- single epoch observations, area spread through the duration of the survey, 465 hours in each of four configuration cycles

In the proposed observing schedule (§ 5.4), we split the ALL-SKY area (minus the area covered by WIDE, leaving 23885 square degrees) into four separate cycles covering the whole area once (476 hours per cycle). An alternative schedule is split into six cycles (1908 hours total, 318 hours per cycle).

The total scheduled time allocated to Tier 1 is **1904 hours** in addition to WIDE, including overhead.

Special Considerations: The data will be taken in a single-pass each, near the maximum slew rate practical for OTFM at the allowed maximum data rate. Thus, there is no opportunity for observing at two hour angles for improved snapshot uv coverage. For the continuum observations, the MFS uv-coverage improvement should partly compensate for this in the imaging quality. This will be verified and quantified as part of the test plan.

Use of BnA Hybrid: We propose in this baseline plan that the observations be taken entirely in B-configuration. However, if practical, we would like to observed the southern area ($\delta < -10^\circ$)

using the BnA hybrid to improve the point spread function. This would require observing 1006 total hours in B-configuration, and 898 total hours in BnA. Note that the time is more heavily weighted towards BnA than the relative fractional area due to the compensation for sensitivity loss at low elevations which impacts VLASS observations only in the low-declination region.

Other Options: It is highly desirable for the discovery of transients that the ALL-SKY tier be observed in two epochs. To accommodate this within the 25MB/s allowed data rate, we would need to drop 5-6 spectral windows and observe with dump times in the range 0.31–0.34 seconds. This mode has not been tested for long periods, and would also incur a sensitivity penalty (relative to the 1500MHz assumed) of 7%–17% due to the decreased bandwidth, necessitating more observing time to reach the desired sensitivity. Verification of this option is an item under the test plan.

3.2 Tier 2

Tier 2 covers a total of 13160 square degrees, including 3160 square degrees of the Galactic plane and the Galactic Bulge, plus 10000 square degrees in the region covered by the SDSS-III footprint. The sensitivity required in this tier is a continuum image rms $\sigma_I = 50\,\mu$Jy/beam, to be uniform over the regions covered. Tier 2 observations will also use OTFM to cover these large areas with high observing efficiency.

The total observing time is 2824 hours for WIDE, and 840 hours for GALACTIC (in addition to ALL-SKY). The Tier 2 total observing time in these regions is **3664 hours**, including overhead.

3.2.1 Tier 2 — WIDE

The WIDE area of 10000 square degrees will require 2417 total hours of integration to reach a final depth of 50 μ Jy/beam rms, or 604 hours of integration in each of four epochs. With our assumed 17% overhead for this Tier, our preferred scenario schedules 2824 hours total of observing, or 706 hours per epoch. For simplicity in scheduling the area covered by WIDE will be observed separately from ALL-SKY in this manner.

Summary for Tier 2 WIDE:

- total area of 10000 square degrees, in SDSS-III footprint
- B-configuration
- combined continuum image rms (Stokes I) $\sigma_I = 50\,\mu$Jy/beam
- total integration time required in this footprint is 2417 hours
- total observing time required is 2824 hours, calculated assuming 17% overhead

Our preferred scenario for observing is:

- four epochs with roughly 12-24 month cadence (e.g. 16–18 months following normal configuration cycling)
- single-epoch continuum image rms (Stokes I) $\sigma_I = 100\,\mu$Jy/beam
- per-epoch on-sky time of 604 hours, taking 706.7 hours observing with 17% overhead, rounded to 706 hours per epoch for convenience
- total scheduled time is 2824 hours

Options: An alternate scenario for observing is to observe spread over six epochs:

- six epochs with roughly 12-24 month cadence (e.g. 16–18 months following normal cycling)
- per-epoch on-sky time of 403 hours, taking 472 hours observing (rounded) with 17% overhead
- single-epoch continuum image rms (Stokes I) $\sigma_I = 122\,\mu$Jy/beam
- total scheduled time is 2832 hours in this scenario

3.2.2 Tier 2 — GALACTIC

The Tier 2 Galactic survey area comprises 560 square degrees of the central Galactic Bulge ($-10^\circ < l < 10^\circ$, $-14^\circ < b < 14^\circ$) plus 2600 square degrees in the Galactic Plane from $-20^\circ < l < -10^\circ$ and $10 < l < 260^\circ$ for $-5^\circ < b < 5^\circ$, for a total of 3160 square degrees. This Galactic region includes significant areas that must be observed at low-elevation even around transit, so integration times are scaled by 1.2× to reflect the higher system temperature (§ 2.1.2); this scale factor assumes observations will be scheduled within +-0.75 hour of transit. For 3160 square degrees, it will take a total of 917 hours of integration to reach a uniform $\sigma_I = 50\,\mu$Jy/beam. The Tier 1 B-configuration observations will have integrated on source for 229 hours, leaving an additional 688 hours to be integrated in Tier 2. The GALACTIC Tier 2 time will be entirely in A configuration. Tier 2 GALACTIC requires an additional 840 hours observing, using the 22% overhead assumed for this component.

Summary for Tier 2 GALACTIC:

- total area of 3160 square degrees
- combined Tier 1 and 2 continuum image rms (Stokes I) $\sigma_I = 50\,\mu$Jy/beam
- total Tier 2 Galactic integration time of 917 hours, with 688 hours in addition to Tier 1 (839 hours calculated with 22% overhead, rounded to 840 hours for scheduling convenience)

A scenario for Tier 2 Galactic observing spreads the additional observations between 4 cycles in A configuration:

- four epochs with 12–24 month cadence (e.g. 16–18 months following normal cycling) in A configuration
- in 4 unique epochs, on-sky observing time of 210 hours each
- single-epoch continuum image rms (Stokes I) $\sigma_I = 115\,\mu$Jy/beam in the A-configuration epochs
- single-epoch continuum image rms (Stokes I) $\sigma_I \sim 100\,\mu$Jy/beam in the single B-configuration ALL-SKY epoch
- total scheduled time is 840 hours

Options: Note that an alternative scheduling scenario spreads these observations among 5 A-configurations instead of 4, each of these would get 168 hours and would reach a depth of $\sigma_I = 129\,\mu$Jy/beam per epoch.

3.3 Tier 3: DEEP

The DEEP tier comprises observations of 3 deep fields with multi-wavelength coverage in a total area of 10 square degrees. The fields are: COSMOS (2 square degrees, 10h00m28s, 02d12m21s), ECDFS (4.5 square degrees, 03h32m28s, -27d48m30s), and Elias-N1 (3.5 square degrees, 16h08m44s, 56d26m30s). The cumulative depth for each field will be an image rms (Stokes I) of 1.5 μJy in A-configuration. There is an multiplicative factor of 1.3× applied to the integration time needed, to compensate for its low elevation assuming observations within 2 hours of transit (§ 2.1.2). For COSMOS, there is an existing $2\,\mu Jy$ S-band C plus A-configuration dataset (12B-158, Smolcic et al.). Note that although these fields are all included in Tier 1, and all but ECDFS is in Tier 2, the amount of time in those tiers in those areas ($<$ 1 hour total per square degree) is not significant enough to count in the total, so we treat Tier 3 as independent. If desired, the B-configuration Tier 1 and 2 data can be included in Tier 3 imaging to provide additional sensitivity to extended structure. These smaller fields can be efficiently mapped using pointed mosaics, so OTF need not be used for the deep fields unless it is desirable to cover these multiple times in a single day for more uniform uv-coverage.

The total scheduled observing time in Tier 3 is 300 hours for COSMOS, 1960 hours for ECDFS, 1131 hours for Elias-N1, which totals **3391 hours**.

3.3.1 Tier 3 DEEP — COSMOS

The summary statistics for Tier 3 COSMOS are:

- total area of 2 square degrees, centered at 10h00m28s, +02d12m21s
- previous observations in 2013-14 (12B-158) to depth of 2 μJy rms equivalent to 302 hours of integration
- combined continuum image rms (Stokes I) $\sigma_I = 1.5\,\mu$Jy (added to previous data from 2012-14)
- total Tier 3 on-sky time of 234 hours (clock time 292 hours with 25% overhead)
- combined continuum image rms (Stokes I) $\sigma_I = 2.27\,\mu$Jy (this data only)

We can observe for 6 hours centered near transit without suffering elevation induced sensitivity losses, so a plausible scenario is to break the observations into 7.5 hour blocks (25% overhead). For 300 hours total, we need 40 blocks. To spread the observing evenly for transient science, we observe this in four epochs:

- four A configuration epochs
- each epoch 10 passes of 7.5 hours observing each (6 hours each integration time per pass, 60 hours integration per epoch)
- single-epoch continuum image rms (Stokes I) $\sigma_I = 4.5\,\mu$Jy
- single-pass continuum image rms (Stokes I) $\sigma_I = 10\,\mu$Jy
- total scheduled time 300 hours (75 per epoch)

Options: Alternative scenarios include spreading the 300 hours among 5 A-configuration cycles (60 hours in each).

3.3.2 Tier 3 DEEP — ECDFS

The summary statistics for Tier 3 ECDFS are:

- total area of 4.5 square degrees, centered at 03h32m28s, -27d48m30s
- combined continuum image rms (Stokes I) $\sigma_I = 1.5\,\mu$Jy
- this would normally take 1208 hours of integration at optimal elevation
- if observed within 2 hours of transit, extra overhead of 1.3× to reach stated sensitivity for elevation effect (§ 2.1.2)
- total Tier 3 on-sky integration time of 1570 hours (1962 hours observing calculated using 25% overhead, rounded to 1960 hours for scheduling convenience)

Since observations should occur within 2 hours of transit, we break the observations of this field into 5 hour blocks (each should get 4 hours integration, 25% overhead, see § 2.3). We will need 392 such blocks total. A plausible scenario is to spread these over four cycles of A configuration epochs.

- four A configuration epochs
- 98 passes of 5 hours each (490 hours) per epoch
- single-epoch continuum image rms (Stokes I) $\sigma_I = 3\,\mu$Jy
- single-pass continuum image rms (Stokes I) $\sigma_I = 29.7\,\mu$Jy
- total scheduled time is 1960 hours

Options: An alternative scenario is to spread the observations between 5 epochs, with two getting 79 passes (395 hours) and three getting 78 passes (390 hours).

Note that by observing closer to transit (e.g. within 1.5 hours) you could decrease the elevation losses incurring less overhead (1.28× increase). However, 3.75 hour blocks would require 128 passes per epoch (over 4 epochs) which is more than 4 months with observations every day. This is likely to have serious logistical problems in scheduling, and thus we choose to take an extra 2% loss in efficiency to observe longer blocks and get back around 8% in overhead which is a net win. If 98 passes are too many, one could observe 2.5 hours from transit at 1.33× penalty (2008 hours total) and take only 80 passes per epoch.

We note that the efficiency loss due to the low elevation has been estimated from a plot of system temperature versus antenna elevation angle at various frequencies in S-band (courtesy R. Perley). Our estimate is uncertain at the few percent level, and could be either too lenient or too conservative. Further testing on the VLA through a trial observation of the ECDFS (e.g. a single pass) will be scheduled as soon as is practical (best done in the upcoming C-configuration or wider).

3.3.3 Tier 3 DEEP — Elias-N1:

The summary statistics for Tier 3 Elias-N1 are:

- total area of 3.5 square degrees, centered at 16h08m44s, 56d26m30s
- combined continuum image rms (Stokes I) $\sigma_I = 1.5\,\mu$Jy
- total Tier 3 on-sky time of 937 hours (1143 hours observing calculated for 22% overhead, rounded to 1131 hours for scheduling convenience)

This field can be observed for long blocks due to its high declination. However, we restrict to around 6 hour integration per block so that they better fit in with dynamic scheduling (but can be observed over a wide range of start times). These blocks will be 7.25 hours long with overhead (in this case 21%), and thus we need 156 blocks total (giving 936 total hours integration, just under our target) for 1131 total hours. Again we split these into four cycles:

- four A configuration epochs
- each epoch 39 passes of 7.25 hours observing each (282.75 hours per epoch)
- single-epoch continuum image rms (Stokes I) $\sigma_I = 3\,\mu$Jy
- single-pass continuum image rms (Stokes I) $\sigma_I = 18.7\,\mu$Jy
- total scheduled time is 1131 hours

Options: An alternative scenario is to divide the observing among 5 epochs, in which case four epochs gets 31 passes (224.75 hours) and one gets 32 (232 hours).

4 Data Products

The VLASS data products are described below, along with risks and considerations in the production and assessment of these products.

Products are broken into classes of ”Basic” and ”Enhanced”, with the former being simple enough to produce by NRAO’s standard (or soon to be standard) data processing system. Enhanced data products will require domain expertise and/or extra resources, so they will be left for communitiy members to define and produce. Both kinds of products will be curated and served to the public by the NRAO, as described below.

4.1 Basic Data Products

The Basic Data Products (BDP) of the VLASS consist of:

1. raw visibility data
2. calibration data and process to generate calibration products (current best version as well as past released versions maintained in archive)
3. quick-look continuum images
4. single-epoch images and image cubes
5. single-epoch basic object catalogs
6. cumulative ”static sky” images and image cubes (generated after epoch beyond the first)
7. cumulative ”static sky” basic object catalogs (generated after epoch beyond the first)

The resources for processing, curating, and serving the data products will be provided by NRAO, as described in §4.3. Teams led by NRAO, but including external community members where possible, will carry out the activities required for the processing and Quality Assurance (QA) of the products.

Details of individual BDP are now described.

4.1.1 Raw Visibility Data

The raw visibility data for the VLASS will be stored in the standard VLA archive. These data will be available immediately after observation with no proprietary period. As the VLASS is being observed using standard data rates (25 MB/s maximum) there are no special resources required for the storage and distribution of these data. Data will be downloadable by users from the archive web pages as normal.

Risk: None. These is standard data comprising a volume that is normal for the amount of time spanned by the VLASS.

4.1.2 Calibrated Data

VLASS data will be processed using a modified version of the normal CASA-based VLA cailbration pipeline. By the time of the VLASS observations, the VLA pipeline will have the requisite functionality to process VLASS data (full polarization, many individual target fields generated in OTF mode). The VLA pipeline is currently run on all VLA observations and is well tested. The VLASS will use a version specifically tested on VLASS pilot observations. See the description of the Calibration Pipeline below for more details.

In summary, the output of the pipeline will consist of:

- a set of final calibration tables
- (possibly) a set of sky models used in calibration
- one or more sets of flagging commands
- the final flag column
- QA reports and plots
- the set of pipeline control instructions or script necessary to calibrate the raw data

Once the pipeline has run and after QA assessment, users will gain access to the calibrated data from the archive. Physically archiving the calibrated visibility data for long-term storage would double the amount of archive space required, which is too costly. Instead, the VLASS will archive calibration products and maintain scripts to generate the calibrated dataset. Scripts to apply calibration products will be based on CASA. Upon first processing of a Scheduling Block the calibrated data will be available for a short time (one to two weeks) for transfer to subsequent processing teams.

Risk: Low. The VLA Calibration Pipeline is in regular current operation, and its code is downloadable by users. We are currently upgrading this pipeline to use the infrastructure developed for the ALMA pipeline. Thus, by the time of the VLASS, this pipeline will have been in regular use and extensively tested. The addition of polarization calibration capability to the pipeline is the only significant addition, and this is expected to be available in the next year also.

4.1.3 Quick-Look Images

The identification of transient and variable objects is a key science goal of the VLASS. This requires the capability to process and image the data within a short period after the data is taken. We have set the requirement on this to be 48 hours maximum, with a goal of 24 hours. If possible we will use the standard VLA Calibration Pipeline described above in a streamlined mode (highly parallelized where appropriate). As a fallback we can use a separate pipeline based on different software, such as the AIPS-LITE based pipeline developed and used by the Caltech group (PI:

Hallinan) used for the S-band Stripe-82 observations taken and processed by that group (VLA Project 13B-370).

The primary output of the Quick-Look Calibration and Imaging Pipeline (QLP) will be continuum images (Stokes I) that can be used by the Transient Object identification teams (see below under Enhanced Data Products) to produce alerts for transient objects and arrange for follow-up studies. As these images will be constructed through mosaicking, there will also be a corresponding rms sensitivity (noise) image.

Of secondary (but still high-utility) priority are the polarized continuum images (Stokes Q,U,V). The Q and U images will be most important for linear polarization studies. The V image is mainly useful as a cross-check on systematics, but may provide useful information for some classes of transients if our imaging and calibration pipeline can produce high-quality V images. We see no reason at this point to preclude the production of V images at this time. There will need to be a noise image for each of these also.

There are thus a total of 8 images to potentially be produced by the QLP per trigger: 4 images (I, Q, U, V), and their corresponding noise maps. As detailed in § 9 these images take a substantial (350TB) but not prohibitive amount of space. The amount needed after a given observation day to search for transients and generate alerts is much smaller. It will be most practical to keep only a buffer of current and some number of past image sets from the same epoch, and only archive long-term the QLP instructions to re-generate previous epochs QL images (most likely in a ”postage stamp” limited area) by ”processing on-demand” (POD). This is particularly true for QL products as these will be streamlined and fast pipelines. We will assume in the resourcing of QL storage that this will be the case.

To manage the computational load, we do not plan to produce the higher-order MFS term images for spectral index (tt1) or curvature (tt2) per epoch. These could be made available as EDP, or as a replacement for the coarse-resolution cubes described below.

We do not plan to produce full channel resolution image cubes in the QL pipeline as a BDP due to the high data volumes possible. This could be part of a EDP proposal, most likely in a POD service offsite.

If possible, for QA purposes or other diagnostic purposes, we will produce coarse image cubes (one plane per 128 MHz spectral window) for IQUV and their uncertainties (8 cubes). Ignoring the 2 or more entire spectral windows lost to RFI (at least the bottom two), there are at most 14 cube planes in each image. Again, these will only be saved for a small amount of time for transient analysis, and if made temporarily available in the archive, no long-term storage will be provided by NRAO (but is a possible EDS capability).

These Quick-Look images will be publicly available from the VLASS archive promptly upon production.

Risk: Medium Production of Quick-Look images will require prompt execution of a Calibration Pipeline, and if the main pipeline (see above) is not fast enough a separate one will need to be developed. Personnel will be needed to oversee running of the QLP and monitor this output, including astronomers and data analysts. Data storage needs are modest and a POD model for past epochs and cubes (under EDP) is practical.

4.1.4 Single-Epoch Images

Within 2 months (Tiers 1 and 2) or 6 months (Tier 3) of completion of the observations of a given ”epoch” (observations in a given configuration), fully calibrated and quality-assured images as detailed below will be produced and available in the archive. These will be generated using a specialized CASA-based Imaging Pipeline. This pipeline will be developed by the NRAO staff, with the involvement and guidance of the VLASS teams.

The VLASS single-epoch wide-band continuum images will include:

1. Flux density calibrated beam-corrected Stokes IQUV continuum (band averaged) images covering the full mosaic area
2. Sensitivity (rms noise) images for the IQUV continuum images
3. Spectral Index and uncertainty images for Stokes I (generated using Multi-Frequency Synthesis)

There are thus 10 single-epoch continuum image products.

Although the wide-band MFS imaging can produce spectral index (Ialpha) images, we suggest that instead of alpha, that the MFS "tt1" (Taylor-term order 1) image be natively stored, as this is effectively I*alpha and will not have the zeroes that come from dividing by I in noisy regions. The archive server can produce a masked alpha image from the native I and tt1 images.

In addition to continuum images, there are potentially cubes of 1024 channels at full spectral resolution for each of:

1. Flux density calibrated beam-corrected Stokes IQUV continuum per-channel image cubes covering the full mosaic area
2. Sensitivity (rms noise) images for each of the cube images

There are thus 8 single-epoch image cube products.

These will contain a wealth of information on the SED and polarization necessary for deriving RM and other computed products. In addition, these will be important for QA purposes and diagnosing RFI problems. Thus they will need to be generated with the processing of a given epoch. However, we note that it will be expensive to store these long-term in the archive for each epoch for Tiers 1 and 2 at least. It is thus attractive to consider a POD on-the-fly processing service for these cubes, as described above for the QL continuum images (§ 4.1.3), hosted on-site or externally (e.g. through XSEDE). It will be of great benefit if this capability is available for Tiers 1 and 2. In addition to efficiency, POD imaging will provide flexibility for choosing coarser frequency resolution in cubes, and the ability to control whether beam-corrections are done or not. The single-epoch Tier 3 images are sufficiently deep and costly that we will assume that cubes are created once and stored in the archive. It is likely that the channels in two or more entire spectral windows with 64 channels each (e.g. the first two at the bottom of the band) will be entirely flagged, and so we assume we need cubes for only 896 channels.

We assume that for Tiers 1 and 2 only modest short-term storage is required for the large full-resolution cubes for QA purposes. For Tier 3 we will archive all of the per-epoch cubes. We will budget for storage of 896 channels in those cases. If storage is a concern, as fallback we will reduce the spatial and/or spectral resolution in these cubes.

For all tiers, it is practical and useful to produce and archive coarse cubes where planes are imaged for each of the 16 spectral windows (64 channels, 128MHz bandwidth) in the dataset. In addition to coarse SED and RM information, these cubes can provide QA diagnoses of RFI or other frequency depended errors. Those would be for the same 9 quantities described above for the full-resolution cubes. Note that it is likely that due to RFI 2 or more of these spectral window planes will be fully flagged and omitted from these cubes. We assume these will require storage for 14 planes.

For all tiers we plan to archive and serve the 14-plane coarse cubes per epoch. If storage is a concern, these would be compressed or made at reduced spatial resolution.

Risk: Low to Medium. Currently, OTF imaging is done using special CASA scripts, with ongoing testing in development code. By the time of the VLASS, it is expected that all needed capability to produce the per-epoch images will be available and tested in a standard CASA release. The fallback is to continue to use development code. Processing for shallow to medium deep observations in Tiers 1 and 2 are of low risk. Single epoch images for the Deep Tier 3 are of medium risk. Development of a robust POD service for Tiers 1 and 2 cubes is of medium risk, with a fallback to long-term archiving of the cubes. Storage of these images is of medium risk, with a number of fallbacks available using compression or reduced resolution.

4.1.5 Single-Epoch Basic Object Catalogs

There are many object finders used in the community for identification and classification of objects from images. It is expected from surveys such as the VLASS that basic object catalogs be produced and released along with the images. Over the next year, we will carry out a study of the available object finders suitable for use on the VLASS data and select and test one (or more if necessary) for production use by the team.

We consider the Basic Object Catalog entry for an ”object” to contain:

1. Position, and uncertainty (likely centroid of I emission)
2. Peak Flux Density (continuum) in IQUV, and uncertainty
3. Spectral Index at Peak (Stokes I) and uncertainty
4. Integrated Flux Density (continuum) in IQUV, and uncertainty
5. Integrated Spectral Index (Stokes I) and uncertainty
6. Basic Shape information IQUV (TBD)

Risk: Low. This has minimal requirements, and it is likely that object finder tools used for previous surveys such as NVSS, FIRST, and the Stripe-82 surveys will be sufficient, if not ideal. It is likely that the studies will identify a superior tool.

4.1.6 Cumulative Images

Within 6 months of the completion of observations of each epoch after the first (1 year for Tier 3), continuum images for the cumulative data will be made. These will be the same set of images as described above for the single epochs. Accommodation for variable objects in the imaging will be required across the multiple epochs.

The VLASS cumulative wide-band continuum images will include:

1. Flux density calibrated beam-corrected Stokes IQUV continuum (band averaged) images covering the full mosaic area
2. Sensitivity (rms noise) images for the IQUV continuum images
3. Spectral Index and uncertainty images for Stokes I (generated using Multi-Frequency Synthesis)
4. Spectral Curvature and uncertainty images for Stokes I (generated using Multi-Frequency Synthesis)

There are thus 12 cumulative continuum image products. These are the images produced for the individual epochs (§ 4.1.4), with the addition of the next Taylor Term (tt2) in the MFS expansion for spectral curvature, plus its uncertainty.

In addition to continuum images, there are cubes of 1024 channels for each of:

1. Flux density calibrated beam-corrected Stokes IQUV continuum per-channel image cubes covering the full mosaic area

2. Sensitivity (rms noise) images for each of the cube images

There are thus 8 cumulative image cube products.

The same considerations apply to the cumulative image cubes as for the per-epoch cubes (see § 4.1.4). The archive will store and serve only the most recent cumulative image cube for each Tier. If the POD capability is available, users can request images from previous cumulative epochs (or possibly other selections). Note that the storage needs for these full-resolution cubes is considerable, and fallback options to reduced resolution or compressed cubes will be necessary to implement in the absence of enhanced archive resources.

It is practical also to create and store the spectral window averaged coarse image cubes (14 planes, 128MHz resolution) at little extra cost. There are 8 of these also. Only the most recent cumulative coarse cube image will be retained.

Risk: Low to High. Tier 1 will be only slightly deeper than its single epoch images and thus low risk. Tier 2 has medium risk, particularly in the confused Galactic region. Tier 3 image production will be challenging (high risk), and extensive algorithm development will need to take place in order to make automated deep imaging (such as that currently manually done on deep fields) work with larger mosaics and higher dynamic ranges required for the Deep tier. This could cause delays in production of final images. Development of a robust POD service for the cumulative data is highly desirable, but not critical as only the most recent cumulative images are of high scientific priority. Data storage for full-resolution cubes is of high (resource) risk. Fallback to reduced resolution may have to be implemented in case additional archive resources (e.g. through EDP) are not identified.

4.1.7 Cumulative Basic Object Catalogs

As in the case of the single-epoch catalogs, these basic catalogs would be made available along with the cumulative images. These are largely based upon the continuum images, we do not foresee using the cubes for these.

We consider the Basic Object Catalog entry for an ”object” in the cumulative catalogs to contain:

1. Position, and uncertainty (likely centroid of I emission)
2. Peak Flux Density (continuum) in IQUV, and uncertainty
3. Spectral Index at Peak (Stokes I) and uncertainty
4. Spectral Curvature at Peak (Stokes I) and uncertainty
5. Integrated Flux Density (continuum) in IQUV, and uncertainty
6. Integrated Spectral Index (Stokes I) and uncertainty
7. Basic Shape information IQUV (TBD)

Risk: Low to Medium. The main risk is in the imaging. All-sky and Wide-Extragalactic are low risk. Will need to handle increased object density and complexity for the deeper tiers and Galactic regions (medium risk).

4.2 Enhanced Data Products

The Enhanced Data Products (EDP) are those that require more domain expertise, and so will be defined and produced by the VLASS community outside the NRAO. These data products will require external support to define, produce, and validate. However, these products are seen as essential to the VLASS science case, so both BDP and EDP will be curated and served by the NRAO.

The initial list of EDP are:

1. Transient Object Catalogs and Alerts
2. Rotation Measure Images and Catalogs
3. Improved Object Catalogs
4. Light curves (intensity and polarization) for objects and/or image cutouts
5. Catalogs of multiwavelength associations to VLASS sources

Beyond requiring extensive domain expertise, these areas are also ideal for nucleating multi-wavelength community groups and resources to work with and enhance the VLASS.

A special case of an EDP is the support for commensal observing at P-band (230–470MHz) using the VLITE system. There is no NRAO processing or archive support currently budgeted for VLITE data products, and thus use of VLITE with VLASS should be considered as an EDP (and EDS) provided in partnership with NRL.

Other areas for EDP will undoubtedly become apparent. Once the survey is approved, we will take proposals for new EDP to be included in the list above. Criteria for including EDP in the VLASS archive will be relevance to the VLASS science case and cost of curating and serving the products. As an incentive to include EDP in the VLASS archive, we ask that the NRAO encourage authors that use EDP to acknowledge groups that produced them.

4.3 Enhanced Data Services and the VLASS Archive

A comprehensive survey like the VLASS will produce a diverse set of data products and will require a full-featured archive to serve it to the public. A baseline plan is to serve products from a website hosted by the NRAO. This site will feature basic search capabilities of catalogs and products, as has been done for FIRST[8] and NVSS[9]. Previous VLA surveys only provided catalogs and images, so at a minimum the VLASS will extend that search capability to visibility data, calibration products, and deep/multi-epoch images. The NRAO will also provide data analysis scripts to apply calibration to raw data.

However, astronomy is increasingly a multi-wavelength discipline with a diverse set of tools for comparing observations from different observatories. If the VLASS archive exists only as a stand-alone NRAO-hosted service, it would not be as useful as one integrated with the tools available at places like IPAC[10] or the Virtual Observatory [11].

We are investigating options for having catalogs and/or images served by organizations outside the NRAO. This would extend the reach of the VLASS outside the radio community and open access to powerful tools for multi-wavelength analysis. The VLASS community, including the co-authors of the VLASS proposal, will be writing an NSF proposal to support VLASS data analysis and a more effective archive. These Enhanced Data Services will greatly augment the utility of the VLASS and its basic and enhanced data products to the wider astronomical community.

Another area that would be greatly improved through EDS is the capability for "processing on-demand" (POD) of images or image cubes. This would alleviate storage volume concerns, and enable more flexible angular and spectral resolution of the resulting products. We expect to utilize

---

[8] `http://sundog.stsci.edu/cgi-bin/searchfirst`

[9] `http://www.cv.nrao.edu/nvss/postage.shtml`

[10] `http://www.ipac.caltech.edu`

[11] `http://www.us-vo.org`

the NSF XSEDE network for modest use of POD-like processing for the pipeline. Fully enabled POD for VLASS could be carried out through partnerships with NSF supercomputing centers or with DOE science labs. Exploration of these options will commence upon approval by NRAO for VLASS.

As noted above, data archive and distribution support for commensal observing at P-band (230–470MHz) using the VLITE system is not currently budgeted for support by NRAO. Archive serving of VLASS should be considered as an EDS provided in partnership with NRL.

4.4 Data Formats

We now describe the formats that we expect to use and serve as data products from the VLASS.

4.4.1 Visibility Data

The VLA archive holds the raw data in the native SDM format. Users can request CASA MS format data will will be produced from the SDM by the archive server.

Users of VLASS can also request the calibration products (tables, flags, pipeline instructions) that they can use to apply to a raw SDM or MS downloaded from the archive.

Calibrated data will be served in CASA MS format upon request through application of the tables, flags, and pipeline instructions.

4.4.2 Images and Image Cubes

The most common image format in astronomy is FITS. CASA can recognize and use (for most operations) FITS images, although it has its own image format based on CASA tables.

The VLASS data services will need to provide the capability to serve FITS and CASA format images upon request for sub-areas (postage stamps, simple rectangular regions) from the available images.

The question of a ”native” format in which to store images is more complex. The most straightforward approach would be to store a set of FITS (or CASA) images that tile the survey areas. Users could request sets of these directly, or ask for a region that would be assembled by the data server to one or more requested FITS or CASA images.

It might be advantageous to take more forward-looking approach and store the image data in a hierarchical data format such as HDF5 (e.g. in the context of LOFAR, Anderson et al. 2010, arXiv:1012.2266). For the all-sky Tier 1, representation in HEALPIX might be useful also. Investigation of options such as these should occur in the year leading up to the commencement of VLASS.

4.4.3 Catalogs

Catalogs could be stored in some internal manner using flat ASCII files, in a relational database (RDB), or in a hierarchical data format (such as HDF5).

Users must be able to get catalogs in simple formats such as flat ASCII files, XML files, or other basic formats.

4.4.4 Plots

Where the data services have the capability of providing plots, these should be in standard formats such as PNG, JPEG, GIF, or even FITS in some cases.

5 Observing

In order to carry out the VLASS, we will need to observe the sky using a large number of mosaicked pointings of the VLA. At 2–4 GHz, the VLA has a field-of-view given by the primary beam response of the 25-meter diameter antennas. This approximately follows a Gaussian response, with a full-width at half-maximum (FWHM) given by

$$\theta_{FWHM} \approx 45' \left(\frac{1\, GHz}{\nu} \right) \tag{2}$$

at observing frequency ν, and thus over the S-band the FWHM varies from 22.5′ at 2 GHz to 11.25′ at 4 GHz, with FWHM of 15′ at 3 GHz mid-band. In order to optimally cover a given sky area in an efficient manner, the array must either conduct a raster scan using "on-the-fly" mosaicking (OTFM), or tile the area with a number of discrete pointings in a hexagonal packed configuration ("Hex-pattern Mosaicking"). The choice between these techniques is determined by the extent to which the extra overhead (from 3 to 7 seconds) needed to move the array and settle at each pointing in the discrete hex-pattern mosaic becomes a burden on the observations, and thus OTFM is favored.

5.1 Mosaicking

The techniques of OTFM and Hex-pattern Mosaicking and the calculations and procedures needed to set these up are described in the Guide to VLA Observing: Mosaicking[1] section. The salient features are:

OTFM: There is very little move-and-settle overhead as the array is in continuous motion over a row of a raster with only a small startup (∼10–15 sec) at the start of a row. In OTFM the phase center of the array is discretely stepped on timescales of a few seconds or longer, so no phase smearing of the images results. However, because the primary beam response pattern is moving with respect to the sky, there are errors introduced in the amplitudes by the moving beam in a single visibility integration time. Thus, the main cost of OTFM therefore is that for fast scanning rates the fundamental integration ("dump") times in the dataset must be short (10% or less) compared to time it takes to cross the FWHM of the primary beam. This in turn increases data rates from those otherwise required (e.g. to avoid time-smearing of the loci in the uv-plane). The secondary cost of OTFM occurs in the imaging process, where the effects of the moving primary beam over the time at which the phase center is fixed must be compensated for by the imaging algorithm at signficantly increased computational cost over that required for a similar observation taken with a fixed pointing center. This is currently done by the CASA software package in its `clean` deconvolution task. Testing of the effacy and efficiency of the use of the CASA imaging for OTFM is underway, and is part of the testing plan given below. For the purposes of this plan, we will assume that OTFM datasets can be imaged with sufficient accuracy to be practical for at least Tier 1 and 2 observations.

Hex-pattern Mosaicking: It takes the VLA 3–7 seconds (depending on the direction of motion in azimuth-elevation coordinates, usually around 6–7 seconds if not optimized) to move and settle between nearby pointings. Thus, if you are spending less than 28 seconds integrating on each pointing the overhead from this motion itself is 25% or higher in the worst case. For a hex-pattern, each field gets 67% of the total integration time desired on-sky, so observations where the VLA Exposure Calculator[12] indicates an on-source time of 42 seconds or less will incur signifcant overhead if not done with OTFM. For VLA S-band, the calculated exposure time is 7.7 seconds at a rms image sensitivity of 100 μ Jy, and thus observations desiring a depth shallower than around

---

[12] `https://science.nrao.edu/facilities/vla/docs/manuals/propvla/determining/source`

$43\,\mu$ Jy in a single pass will prefer OTFM. Thus, our Tier 1 and 2 observations with single-pass depths $> 100\,\mu$ Jy will require OTFM, while the deep Tier 3 observations can be carried out with straightforward Hex-pattern Mosaicking. Note that it is in principle possible to arrange the order in which the fields are observed to make the motions to be predominantly in the elevation axis which damps the telescope settling, and has been shown to lose only 3–4 seconds. However, this optimization is dependent on the local time at which the observations are made requiring restricted LST scheduling blocks, and is difficult to arrange over large areas of the sky. We will try where possible to use this optimization to further reduce overheads in the Tier 3 mosaics.

Risks: Low to Medium Other than increased overheads, the Hex-pattern Mosaicking for Tier 3 does not incur signficant risk. OTFM has been used successfully for shallow (single-pass image rms $\sim 100\,\mu$ Jy) in S-band in the Stripe-82 observations by Hallinan et al. (program 13B-370). Images with rms $\sim 60\,\mu$ Jy from the three passes spanning 2 months are being made as part of the testing program, and current indications are that they are of acceptable quality. Thus, for Tiers 1 and 2, we deem the risk using OTFM to be low overall on the basis of quality. The further consideration is the extent to which more expensive imaging costs are required in areas in which there are bright sources, particularly for polarimetry, and for accurate determination of fainter source spectral indices. This is also being assessed in the testing. As the new imaging algorithms are only now available (July 2014), we assign this as medium risk to resourcing for the VLASS, as extra computational power or longer processing times are likely to be required in some regions.

5.2 Scheduling Considerations

The Jansky VLA is normally operated using a ”Dynamic Scheduling Queue” where the individual Scheduling Blocks (SBs) are created in the VLA Observing Preparation Tool (OPT) to be able to be executed in a prescribed range of LST, and submitted to the VLA Scheduler software (OST) to be queued up for observation by the array in a manner dictated by weather and priority. It is our intent that the Tier 3, along with some of the Galactic Plane Tier 2) observations, where possible for each pass, be observable using standard Dynamic Scheduling. This requires that these blocks contain sufficient calibration to stand alone or be boot-strapped from other VLASS SBs executed nearby in time.

For Tier 1 and most of Tier 2, it would be advantageous to construct the schedules in large blocks to be observed at specific LST start times. This would allow maximum efficiency in calibration and control of slewing (e.g. telescope wraps). In practice, due to considerations such as the ability to allow interrupts for target-of-opportunity observations, and fault tolerance (e.g. for power outages, weather, etc.), the schedule will need to be broken in to modestly sized blocks. Ideally, the VLA Scheduler software would be able to handle sets of SBs that are linked (e.g. in a particular order or in alternate sets). However, this capability does not currently exist, and there may not be resources in the software group to allow this. Instead, the most straightforward plan is to make sets of SBs for submission (see below) and submit only a day ahead. This will require an ”Astronomer on Duty” (AoD) for VLASS who will keep track of what SBs are ready to observe, make sure they are submitted, make sure that they run, and make any modification necessary (e.g. due to TOO or weather interrupts).

Target-of-Opportunity Interrupts: For the long-block observations in Tiers 1 and 2 (and some of Tier 3), provision will be made for the possibility that observations will be interrupted for time critical TOO programs (e.g. for triggered transient observations). There is currently no mechanical provision in the way schedules are constructed or executed for the supension and restarting of schedule blocks. Therefore, the most straightforward implementation is to break all schedules into blocks of 2–3 hours in length, and to allow TOO interrupts to simply stop the execution of the current schedule and possibly pre-empt the execution of the following one or more SBs. After TOO observations are complete, the VLASS schedule would resume with the next appropriate block. The AoD would be informed of this interruption, and would examine the archive record to determine the missing observations and construct a ″make-up″ SB to be run at the first appropriate opportunity. The plan for the construction of schedules (see below) will take this need into account.

Risk: Medium The VLASS scheduling and observation monitoring will be a vast book-keeping exercise. As described above we believe this is controllable, either through some modest improvements in operations software, or at worst with some workarounds in the way we carry out the scheduling. There will need to be a pool of VLASS astronomers and data analysts to fulfill AoD staffing throughout the survey. This ideally might include students participating in the VLASS from the community. Some NRAO resourcing will be required to support these activites both directly and for supervision.

5.3 Schedule Construction

It will be impractical to use the VLA OPT in standard interactive mode to construct the thousands of hours of schedules for thousands of pointings that must be in the Source Catalog Tool. Instead, we will create some lightweight software in Python to construct the ascii lists that the OPT and SCT can read. This feature has been used by the 13B-370 Stripe-82 observers (mainly Caltech graduate student Kunal Mooley) with success to schedule those observations. Our plan is to use and modify as needed the Python code developed by Kunal for that project. These scripts and code would be made available to the community for their own use for similar surveys.

Risk: Low We can build upon the process used by 13B-370 for the Stripe-82 survey for schedule and catalog construction. Some rules for keeping track of schedules and archiving will need to be devised and followed.

5.4 Overall Observing Schedule

The VLASS as proposed will be carried out over the course of at least 4 cycles of JVLA in its A and B configurations, spanning a total of 5–6 years (50–72 months or more). We present here two possible schedules for observing the VLASS. Other scenarios are possible, should new constraints be imposed (e.g. based on design reviews).

5.4.1 Scenario I: Proposed Observing Schedule

The first scenario is a "5-year" survey spanning 4 cycles and a mininum of 56 months:

Cycle	Config	ALL-SKY	WIDE	GALACTIC	COSMOS	ECDFS	Elias-N1	Total
1	B	476	706	0	0	0	0	1182.0
1	A	0	0	210	75	490	282.75	1057.75
2	B	476	706	0	0	0	0	1182.0
2	A	0	0	210	75	490	282.75	1057.75
3	B	476	706	0	0	0	0	1182.0
3	A	0	0	210	75	490	282.75	1057.75
4	B	476	706	0	0	0	0	1182.0
4	A	0	0	210	75	490	282.75	1057.75
Total		1904	2824	840	300	1960	1131	8959.0

Note: The ALL-SKY allocation is placed under B configuration, while in practice we we would prefer to observe the southernmost regions in the BnA hybrid to form a more circular synthesized beam for imaging, comparable to that acheived at higher declinations (see § 3.1). If this option is used, then 1006 hours would fall in B-configuration (251.5 hours per cycle) and 898 hours would be observed in BnA (224.5 hours per cycle).

This starts in May 2016 and follows the standard cycle schedule (BADC), completing at the end of 2020 (or early 2021). We assume that each configuration is the current 4-month duration, but note that accommodating the VLASS could involve extending some of the durations in the more loaded sessions. The B-configuration ALL-SKY (excluding the WIDE area) has a single epoch per field, with the fields distributed across four cycles. This was the scenario presented in the main proposal document.

5.4.2 Scenario II: Alternative 7-Year Schedule

In the following alternative scenario, we spread the VLASS over 7 years and parts of 6 cycles (84 months minimum):

Cycle	Config	ALL-SKY	WIDE	GALACTIC	COSMOS	ECDFS	Elias-N1	Total
1	B	318	472	0	0	0	0	790.0
1	A	0	0	168	60	395	224.75	847.75
2	B	318	472	0	0	0	0	790.0
2	A	0	0	168	60	395	224.75	847.75
3	B	318	472	0	0	0	0	790.0
3	A	0	0	168	60	390	224.75	842.75
4	B	318	472	0	0	0	0	790.0
4	A	0	0	168	60	390	224.75	842.75
5	B	318	472	0	0	0	0	790.0
5	A	0	0	168	60	390	232	850.0
6	B	318	472	0	0	0	0	790.0
Total		1908	2832	840	300	1960	1131	8971.0

In this longer scenario, the survey starts with B configuration in 2016A in May 2016, and completes in 2023. It spreads ALL-SKY and WIDE between 6 B-configuration cycles, and the DEEP and GALACTIC among 5 A-configurations, impacting each cycle less heavily. Each cycle gets no more than 850 hours (35.4 days) of scheduled VLASS.

5.5 Schedule Pressure by LST

Using the primary schedule (§ 5.4.1) for 8959 hours spanning 4 configuration cycles over 5 years, we have estimated the total LST pressures in the B and A configurations for the various components. These pressures represent the average number of ″passes″ at a given LST needed to carry out the VLASS.

Formally, we approximate each component of the VLASS as a total time T broken into observing blocks that can be observed in a ″window″ of length W starting at some LST H. All times from H to $H+W$ for that component are assigned a pressure $P = T/W$. We estimate the following parameters for the VLASS:

Component	Config	T(hrs)	W(hrs)	H	P
ALL-SKY + WIDE(1 epoch)	B	2611.0	24.0	0^h	108.8
WIDE-NGC (3 epochs)	B	1412.0	19.0	3^h	74.3
WIDE-SGC (3 epochs)	B	706.0	13.0	17^h	54.3
GALACTIC Plane	A	656.0	18.0	16^h	36.4
GALACTIC Bulge	A	184.0	5.5	15^h	33.5
DEEP COSMOS	A	300.0	9.0	5.5^h	33.3
DEEP ECDFS	A	1960.0	5.0	1^h	392.0
DEEP E-N1	A	1131.0	12.0	10^h	94.2

For ALL-SKY and WIDE, it is broken into a uniform component (ALL-SKY plus a single epoch of WIDE) and 3 epochs of WIDE. To construct the plot a single epoch of WIDE is subtracted from the uniform combination and added back in to the WIDE 3 epoch values to form the full WIDE and the ALL-SKY excluding WIDE. Note that wide is itself broken into two large patches in the North Galactic Cap (NGC) and South Galactic Cap (SGC) areas at an estimate areal ratio of 2:1 from Figure 10 in the VLASS Proposal (we will get more accurate regions for the actual scheduling).

Plots of the total VLASS LST pressures are presented for B-configuration (Fig 1) and A-configuration (Fig 2).

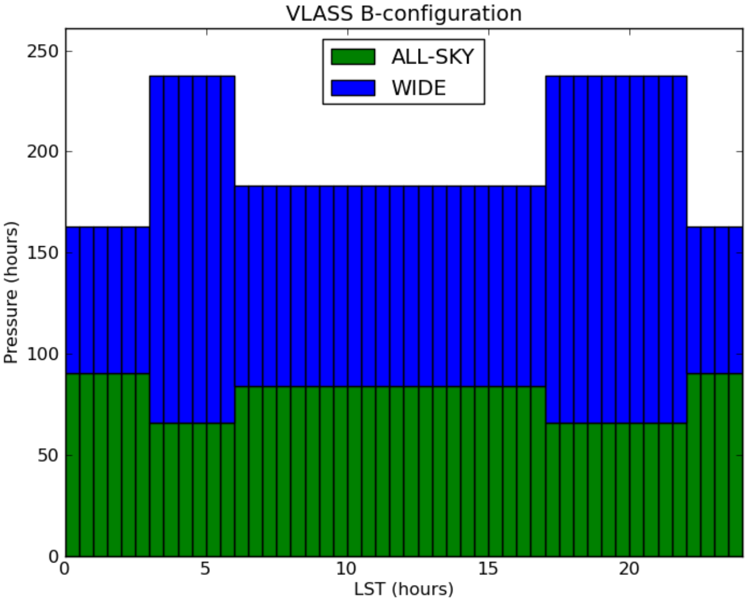

Figure 1: Total pressure by LST for the components of the VLASS observed in B (and possibly BnA) configuration. Hours are calculated for the primary schedule scenario (I) with observations spread through 4 configration cycles spanning 5 years, using the methodology described in § 5.5. The ALL-SKY pressure (green) excludes the area covered in WIDE (blue), with time equivalent to a single epoch of WIDE. The high pressure regions are where the North Galactic Cap (3^h–21^h) and South Galactic Cap (17^h–6^h) windows overlap, the actual pressure could be more evenly distributed by narrowing the windows. Note that use of BnA in ALL-SKY (898 hours) would be an equivalent pressure of 37.4 hours on this plot uniformly spread over all LST.

6 Calibration

The goal of the VLASS Calibration process is to determine, on the basis of *a priori* factors and from observations of standard calibration sources, the corrections to the raw data amplitude, phase, and visibility weights to be applied to the data. This process also determines the flags that are needed to remove bad data due to instrumental faults, RFI, and other causes of error. When applied to the VLASS data, this calibration will allow the production of images in the next processing stage. This process only includes the derivation of the complex gain and bandpass calibration factors known

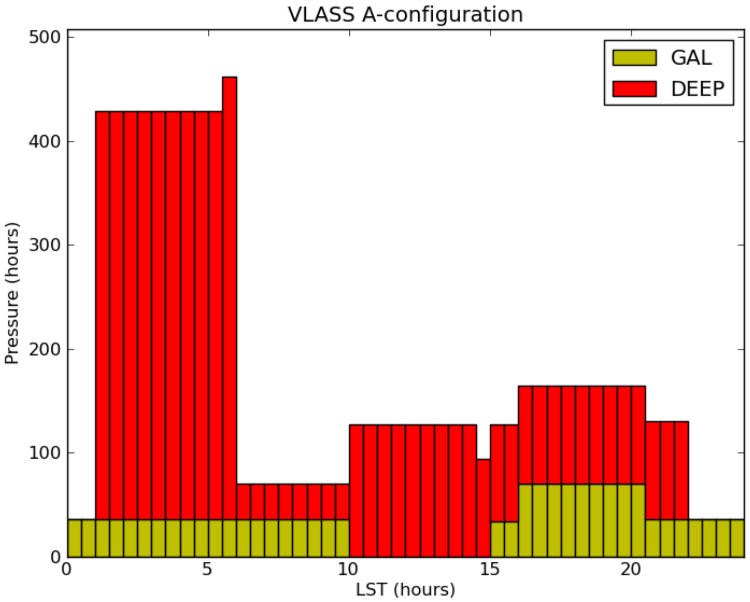

Figure 2: Total pressure by LST for the components of the VLASS observed in the A-configuration. The hours calcuated are for the primary scenario (§ 5.4.1) spread over 4 configuration cycles spanning 5 years. The GALACTIC component pressure shows the peak around 18^{h} for the Bulge. The large peak in the DEEP pressure corresponds to the ECDFS window.

through previous measurements or determined by the observations of calibrators and transferred to the VLASS target observations. The self-calibration of VLASS data is included in the Imaging stage of processing.

VLASS data will be processed using a modified version of the normal CASA-based VLA cailbration pipeline. By the time of the VLASS observations, the VLA pipeline will have the requisite functionality to process VLASS data (full polarization, many individual target fields generated in OTF mode). The VLA pipeline is currently run on most VLA observations and is well tested. The VLASS will use a version specifically tested on VLASS pilot observations.

The VLASS Calibration Pipeline will carry out

1. Application of initial online flags (off-source, focus error, subreflector error)
2. Determination and application of derived flags (RFI, bad antennas, shadowing, other)
3. Switched power amplitude calibration and antenna gain curves
4. Flux scale calibration (using standard sources)
5. Complex Delay and Bandpass Calibration
6. Complex Gain Calibration
7. Flux density bootstrapping (from primary to secondary calibrators)
8. Polarization On-Axis Leakage Calibration

9. Linear Polarization Angle Calibration
10. Interpolation and Application of Cumulative Calibration
11. Final Flagging of Data (insufficient or failed calibration, RFI)
12. Output of Quality Assurance (QA) information, plots, images

As mentioned earlier when discussing Data Products, it is more efficient to store the calibration tables, flags, and pipeline commands and the create the calibrated dataset upon request from the archive, rather than to store both raw and calibrated datasets. Should it be deemed appropriate and possible, we might consider also archiving off-site the full calibrated dataset (e.g. through Enhanced Data Services by a community partner).

Risk: Low to Medium The VLASS is well-suited to pipeline processing and will be able to make use of the VLA Calibration Pipeline as it is currently being developed. This pipeline is now being re-implemented in a new archtecture based on the ALMA pipeline development, and by the time of VLASS will be well tested. Incorporation of polarization in this pipeline will be a primary goal over the coming year, and is necessary not just for VLASS. There are minor medium risk issues in the calibration for the Quick Look images.

6.1 Observations of Calibrators

The first concern for calibration is setting up the observing such that sufficient calibration can be performed on the data. The calibration of the Tier 3 DEEP fields is straightforward, as they require only one or two good calibrators per field.

The calibration of the entirety Tiers 1 and 2 will be more difficult. We will require the availability of a ”network” of suitable calibration sources spread across the entire visible sky with a spacing of one every few degrees — a density of around 0.05–0.25 deg^{-2} giving a list 1700–8500 in total. These will need to be compact (for the VLASS A and B configuration observations). Initially, the obvious starting place is the current VLA Calibrator Database, currently containing 1865 sources. Of particular interest is the VLBA calibrator database, with around 4700 sources in the VLASS area, as these sources are known to be compact and have astrometric precision positions.

Another good starting set is the CLASS catalog of compact sources measured at 8.4GHz in A-configuration Myers et al. (2003). The CLASS database contains over 13000 distinct detected sources in the northern celestial hemisphere, a good fraction of which would be suitable as VLASS calibrators. There are extensions to the VLA-visible southern celestial hemisphere, including the compact calibrator catalog of Winn et al. (2003). These have some overlap with the VLA and VLBA calibrator databases but should yield a substantial number of excellent calibrators.

Prior to starting the VLASS Tiers 1 and 2, it would be prudent to observe candidate calibrators from these lists to identify those suitable for our survey. This could be carried out for 6000 targets with around 60 hours of observations. Ideally, this would be done in S-band B or A configuration, but it may also be practical to carry out at C- or X-band in the C-configuration previous to the start of the VLASS.

Identification of calibrators for the Tier 2 GALACTIC area will be more difficult. Most of our known calibrators come from parent surveys that avoided the galactic plane and bulge. Some additional work will likely to be required to find a good dense set of calibrators in this region. For example, one could contemplate a shallow fast-scanning survey covering the $\sim 3200\,\mathrm{d}eg^2$ area in around 40 hours, sacrificing imaging quality and spectral bandwidth and sensitivity for scan speed within allowed data rates. This would also be of general benefit to VLA users.

Note that once the first epoch of the Tiers 1 and 2 are complete, the survey itself will yield a full (unbiased) list of calibrators for use in subsequent epochs.

Identification and characterization of the VLASS calibrator list is a key item in the Test & Development Plan below (§ 10).

Risk: Low to Medium. There are viable paths to obtaining an initial calibrator database, although ideally verification of this database would require up to 100 hours of test observing prior to the VLASS.

6.2 Calibration Issues for the VLASS

Because the VLASS covers the entire sky usefully visible to the VLA, there are a number of issues related to calibration that require special attention and investigation. These are mostly related to the RFI known to exist at 2–4 GHz, but also include possible ionospheric issues.

6.2.1 RFI at S-band

As described above in § 2.1.1 substantial parts of the 2–4 GHz band are plagued by RFI. Some (though not all) of the most pernicious interference comes from geostationary satellites in the ”Clarke” belt, which appears at Dec $\delta \approx -5.5^\circ$ from the VLA. This will affect VLASS observations near (within 10 degrees from our current estimation) this Declination. In addition to reducing the usable bandwidth, RFI can affect the calibration of the data. Investigation of this effect and the devising of observing and calibration strategies to mitigate RFI induced problems are part of the Test and Development Plan (§ 10.1.1).

Case Study: Stripe-82. The B-configuration S-band observations (12A-371,13B-370) in Stripe-82 encountered calibration issues attributed to RFI. In the first set of observations in 2012 (12A-371) covering 50 square degrees, the sub-stripe was broken into two sections that were calibrated using different calibrators, one slightly north of the stripe (J2212+0152) and the other to the south (J2323-0317). During processing, it was noticed that the observed amplitudes on the southern calibrator J2323-0317 varied greatly as a function of time (azimuth), and that application of the gains derived from these scans to the target data introduced spurious variations in the flux densities and rms variations in off-source areas of the images. The observations of the northern calibrator J2212+0152 showed little or no variation. There did appear to be some variation in the Stripe 82 target observations, and thus the effect was still present on the equator. Some of the hour angles for J2323-0317 most strongly affected appeared to correspond to azimuths of known geostationary satellites, and thus compression in the receiver system due to RFI from these satellites was strongly suspected. When the observations for the full 270 square degrees of Stripe 82 for 13B-370 were carried out, schedules were constructed avoiding the hour angles most strongly impacted in the 12A-371 data. In addition, calibrators to the south of Stripe-82 were not used. The discovery and mitigation of this effect in the 50 square degree sub-stripe are described in more detail in the paper by Mooley et al. 2014 (in preparation). More extensive tests on the full Stripe 82 from 13B-370 are still underway, and will also be part of the VLASS Test and Development Program.

Risk: Medium RFI is clearly going to be a problem near the Clarke belt, and possibly elsewhere. Observations in these affected regions will likeyly need to be treated specially during processing, with extra resources and personnel requirements.

6.2.2 The Ionosphere at S-band

There will probably be some times when ionospheric disturbances will impact the calibration and imaging of the data. Some of these can be avoided by use of dynamic scheduling, or re-observing of blocks damaged beyond repair. Use of pre-calibration (e.g. using the TECOR algorithm) and imaging/self-calibration iterative schemes coupled with global sky models (evolved as the survey progresses) will be employed as needed for the rest of affected data.

Risk: Low to Medium. The extent of this is unknown. Recent observations and reduction of data from Stripe-82 and COSMOS at S-band do not appear to be strongly impacted by the ionosphere, although occasional times requiring careful self-calibration were encountered, and were possibly due to ionospheric issues. The main risk will be in the time and effort required to identify affected data and to re-process using mitigation techniques.

6.3 Algorithm and Software Development for the Calibration Pipeline

Only a modest amount of additional software development is required for the VLA Calibration Pipeline to handle the VLASS data. In particular, accommodation of OTF scans in the pipeline (e.g. supression of flagging at beginning and end of scans for the short OTF scans) is the primary issue. Otherwise, the VLASS data can be treated as any other VLA dataset.

The other issue for use of this standard VLA pipeline for VLASS is its overall speed and efficiency. In particular, use of the pipeline within the Quick Look image production path will require it to process data promptly upon observation. The current VLA Calibration Pipeline is too slow, although bottlenecks have been identified (e.g. the speed of `applycal`, production of QA plots). It is likely that development of a version of the pipeline that streamlines processing for VLASS will be necessary. Exploration of options for speed-up for QL and for general use are part of the Test and Development Program.

Risk: Low to Medium The modest changes in the standard pipeline here should not require significant resources. Use of the pipeline for normal VLASS use is straightforward and low risk. Use for Quick Look processing might be medium risk if our efforts to speed it up are insufficient. Fall-backs include using more cluster nodes for parallel processing of single SBs, or use of the Caltech AIPS-Lite calibration pipeline.

6.4 Calibration Processing Requirements

Estimation of the processing resource requirements required for carrying out the VLASS operations outlined above is fairly straightforward. We now have experience running the current (script-based) JVLA Calibration Pipeline for all observed projects. In addition, the Stripe-82 and COSMOS programs have used custom calibration scripts as well as the standard calibration pipeline. All of these use CASA in the currently available (version 4.2.2 and earlier) standard ”serial” processing mode. This forms the baseline for our processing time estimates.

6.4.1 Case Study: Calibration of Stripe-82 datasets

For the purposes of VLASS testing, we have created a custom crafted CASA calibration script for the 3-hour blocks of the 13B-370 program. This goes end-to-end from the importing of a raw SDM file to a CASA Measurement Set, through to the creation of a calibrated MS containing only the target data. As of this time, it does not include polarization calibration (though this should only increase the processing time by a small factor, perhaps 10%). It also does not create plots equivalent to the standard pipeline ”weblog”.

The test data set was one of the 3-hour blocks from the 13B-370 Stripe-82 program[13]. Running the custom calibration script took 10.8 hours of elapsed time, or a factor of 3.6 times the observing block duration of 3 hours. This implies that it is feasible to run a serial calibration process in less than 4 times the observing time, and thus it is practical to keep up with observing by running 4 independent calibration processes on a single cluster node (the memory footprint for calibration appears to be sufficiently small). For reference, the standard VLA Calibration Pipeline[14] took

---

[13] SB `13B-370.sb28581653.eb28626177.56669.781848645835`

[14] `https://science.nrao.edu/facilities/vla/data-processing/pipeline`

66.25 hours to process this same 3 hour block (22 times the block duration), so some substantial efficiency improvements (including disabling of much of the plotting) would be required to employ this for the VLASS.

The calibration for Quick-Look imaging is the most constrained, as our specification is to fully process the data and make images available within 48 hours of the completion of an observation block, with a goal of 24 hours or less. If the serial processing of a block takes 4 times the block duration, then observing blocks would need to be 6 hours or less in duration to process within 24 hours (leaving 24 hours for imaging to meet the specification), or 3 hours if you want to process within 12 hours (half of the goal of 24 hours total). Implementation of an truly parallel calibration pipeline would allow use of longer observing blocks.

Risk: Low The modest changes in the standard pipeline here should not require significant resources. Use of the pipeline for normal VLASS use is straightforward and low risk. Use for Quick Look processing might be medium risk if our efforts to speed it up are insufficient, although custom scripts are clearly feasible. Implementation of intrinsic parallelization would be ideal and greatly beneficial, but is not required to meet our specifications.

7 Imaging

The VLASS is at its heart a wide-band continuum imaging survey. The science goals of the survey are predicated on the ability of the instrument and data processing to deliver images of sufficient quality to be able to identify objects and measure the salient properties (e.g. flux density, position, spectral index, polarization, light curve). The goal of the VLASS Imaging Pipeline is to

Starting with calibrated visibilities output from the Calibration Pipeline, the Imaging Pipeline must:

1. Select a sub-region of the sky to image
2. Gather the visibility data that are relevant to that sub-region of sky
3. Carry out wide-band continuum imaging for that sub-region
4. Assess the quality of the imaging, determine whether further imaging, self-calibration and/or data flagging iterations are required
5. If required, perform iterative imaging steps
6. Output final images, sky models, and QA information and plots

We now describe the requirements on the VLASS Imaging Pipeline, discuss software development needed to carry this out, and estimate the processing resources that will be used to image the VLASS.

7.1 Imaging Requirements

In order to keep up with the observing, the Imaging Pipeline must be able to process the data at a rate commensurate with the observing rate. This will be effected through the parallel image processing of sub-mosaics on NRAO-based clusters or externally provided systems (e.g. through XSEDE).

There are three imaging processes that need to be handled by the pipeline:

- Quick Look (QL) imaging triggered after every scheduling block is observed (e.g. for transient identification)

- per-epoch imaging triggered after the last observation each configuration
- cumulative imaging triggered after each epoch beyond the first, incorporating all previous data

For each of these, there are three kinds of images that may be produced:

- Wide-band (2–4GHz) continuum images
- Full-resolution (2MHz channels) image cubes
- Coarse-resolution (128MHz spectral windows or similar) image cubes

Imaging is done in all Stokes parameters (IQUV) for polarimetry capability.

Continuum imaging can include higher-order Taylor terms in the spectral dimension (e.g. spectral index, spectral curvature) depending on image depth (e.g. for processing beyond the Quick Look). CASA has algorithms for this that have been used in past programs, and further development of these capabilities is underway. Full-polarimetric imaging is a key part of VLASS, and the use of accurate polarized "primary beam" maps of the VLA field-of-view during imaging and analysis are critical to the production of science ready images. Self-calibration (through the use of previous sky models as well as true self-calibration from iterative imaging) is also an integral part of the image processing.

We assume for all image size calculations that in the ideal case the images will be pixellated at a sampling level 0.4 of the (robust weighted) resolution at the highest frequency of the band (4 GHz), rounded to a convenient value. For A-configuration, this is $0.2''$ (resolution $0.49''$ at 4GHz), giving 324Mpix per square degree. For B-configuration, this is $0.6''$ (resolution $1.58''$ at 4GHz), or 36Mpix per square degree. In practice we will refine this based on imaging performance and we may be able to get by with less oversampling. In addition, individual sub-images will likely need some amount of extra padding to accomodate odd shapes. Overall we should treat these estimates as reasonably conservative, uncertain at the around the 25% level. However, this optimal level of resolution leads to large image archive sizes (see § 9) and thus a key issue for testing is the determination of the lowest acceptable resolution in the images and image cubes that will still enable the key science with the Basic Data Products. It may be possible to reduce the image data volumes by significant factors (2–8) in this manner. In addition, image compression algorithms (lossless and lossy) can also improve storage efficiency and will be investigated in the Test and Development Program.

The image cubes are needed as input to more advanced processing for Rotation Measure determination, spectral line surveys, and more detailed SED modeling of sources. Most of these would be provided as Enhanced Data Products and Services. Note that the storage and distribution of the large full-resolution cubes will be a challenge for the archive (see § 9), and options for external hosting and "on-demand" image processing should be explored (e.g. as an Enhanced Data Service). As a fallback we would carry out compression through a combination of reduced angular resolution (e.g. $0.3''$, Nyquist at 3GHz) and spectral resolution (average 8 channels, 32MHz).

Risk: Medium to High Previous Jansky VLA programs (e.g. Stripe-82 and COSMOS observations) have been imaged using scripts and software tailored to the individual projects. As of this time, there is no standard CASA-based JVLA imaging pipeline available. The development of this capability is a high priority for NRAO in the next two years, and thus the VLASS can build upon and help pioneer this effort. We expect this to take substantial resources (design, development, implementation, execution, and QA), some of which will need to take place also as part of the normal JVLA operational plan. Due to their similarity to Stripe-82 processing, we expect the pipelining of Tier 1 and Tier 2 Extragalactic imaging to be only medium risk, with the possibility of falling back to the processed used as part of 13B-370. We assess the risk for Tier 2 Galactic and Tier 3 Deep Fields to be higher, due to the considerably more stringent dynamic range constraints, the complexity of the emission in the Galactic fields, and the relative lack of previous experience in automated image processing for large deeper mosaics. These risks will be mitigated through the VLASS Test and Development Program (§ 10) using existing data taken in similar modes of observing.

7.2 Algorithm and Software Development for Imaging

The development and implementation of improved software in CASA for imaging is critical for the success of the VLASS. For example, the Quick-Look imaging is the most time-critical, while the DEEP and GALACTIC imaging will be the most difficult from dynamic-range and image complexity perspectives.

The issues to be addressed include:

- the robust parallelization of imaging with minimal memory footprint (so multiple proceses can be run on a single node) and optimal I/O (to keep disk access traffic to only what is required);
- the efficient joint mosaicking over a modest area (4 × 4 primary beams) with multiple Taylor terms, including W-projection (and A-projection), multi-scale, for full polarimetry;
- handling of primary beam effects during imaging (e.g. A-projection) and for correction of the final images ("`wfpbcor`") at the appropriate level and accuracy for the VLASS components;
- the use of pre-boxing (based upon other catalogs or evolving VLASS global sky models) and auto-boxing;
- ability to carry this out in a semi-automated pipeline that includes fault detection and handling, robust and frequent check-pointing of the run to allow restarting of the imaging upon failures or crashes.

Note that these are also general development targets needed to fully realize the potential of the JVLA and ALMA for high-quality imaging. The needs of the VLASS are not special (the DEEP component is not as deep as existing and planned observations) but the large volume of data and the requirement to produce science capable images on our schedule do place constraints on the timescales for this work, and thus require substantial resources for the software development and testing. Naturally, apects of this work are part of the Test & Development Plan (§ 10.2 and 10.3).

7.3 Processing Requirements for Imaging

The imaging of the VLASS data will clearly be the computational bottleneck for processing. However, it is also the most straightforward part of the procesing to run parallel serial scripts on, as individual fields or sub-mosaics can be processed independently and combined together in a final linear mosaic (or left as "postage stamp" sub-images). Careful benchmarking using representative test data for the various VLASS Tiers and fields will be required to obtain accurate estimates of required resources. This will be a critical task under the Test & Development Program, and will require significant effort to carry out.

As an illustrative example, we consider a simple test case based upon existing data.

7.3.1 Test Case: Imaging of Stripe-82

A custom simple imaging script for the same observing block used for the calibration benchmarking (§ 6.4.1) was run alone on a single cluster node. The observations contained in this 3 hour observing block covered approximately 22.5 square degrees of the sky reaching a depth of $\sim 90\,\mu Jy/$beam rms for the full band. On average, it took 145 seconds per field, or 81.56 hours for all 2025 OTF phase center fields. Thus, the basic imaging for our 3 hour observing block took approximately 27 times the block duration. Thus, imaging could keep up with observing by running the imaging of individual fields in parallel using 7 nodes with 4 processes per node.

Note that the script used for this test only did very basic imaging of each OTF phase center as if they were single fields. No joint mosiacking was employed, nor was any self-calibration carried out. This did a single MFS Taylor term and single scale deconvolution, did not include any autoboxing, but did use W-projection to deal with the wide-field aberrations. The cell size employed was $1''$ matched to the band center, rather than the more optimal $0.6''$ proposed for VLASS (§ 9.3), so the VLASS images would have 2.8 times the number of pixels compared to images generated in this test. Our test thus provides a lower limit on the processing required for the Tier 1 ALL-SKY and Tier 2 WIDE imaging, and an approximation of the requirements for Quick-Look imaging in most tiers. Given the above considerations, one might conservatively extrapolate that VLASS imaging will require around 4 times the resources indicated by this test, requiring around 30 nodes (with 4 processes per node) to image in the same time as observing (e.g. 6 hours to image a 6 hour block).

These preliminary estimates carry significant uncertainty, and more careful studies are required to provide accurate projections of resources needed for VLASS imaging. In particular, detailed tests using the deeper COSMOS data and test observations in the Galactic plane, are needed to give reliable estimates for the imaging of the DEEP and GALACTIC components of the survey.

8 Image Analysis and Sky Catalogs

The main image analysis task for the VLASS is the production of the basic object catalogs for the Quick-Look and standard images.

A good study of the performance of radio continuum image source finders is Hancock et al. (2012), which considers the available options in the context of ASKAP. The upshot is that there are options available that should have acceptable performance for the basic catalogs for VLASS. Note that inclusion of the spectral index images and polarimetric images from VLASS will likely require some extensions to these source finders, which in turn will require some developer or astronomer time.

Also available as a proof of concept is the source finding carried out for the JVLA Stripe-82 surveys by Mooley et al. (in preparation). There is also a comprehensive discussion in Mooley et al. (2013) in the analysis of archival VLA ECDFS multi-epoch data. It is our current assessment that one or more of these methods will be suitable for the basic catalogs from the VLASS.

More advanced catalogs and source finding algorithms could be developed and produced as an Enhanced Data Product.

Risk: Low to Medium There are a number of options available which will do some of what we want, if not all. Dealing with the wide-band spectral index and polarimetric data products of the JVLA will likely require some amount of extra work.

9 Archiving and Data Distribution

The primary interface that the user community will have to the VLASS is through the archive and data distribution system. Raw data will be served via the normal JVLA archive, available with no proprietary period as soon as it has been ingested into the archive system.

The archive, or at least some archive, will have to also serve the VLASS data products as described above. It is the responsibility of NRAO and the VLASS to make the Basic Data Products available through this archive mechanism. Enhanced Data Products may be made available through the NRAO-hosted VLASS archive, this will need to be negotiated and is largely dependent upon resources required. The VLASS products, either in basic form or further processes, may also be made available via alternative Enhanced Data Services, as described above.

The estimated ideal data volumes required for storage of the VLASS data and data products are:

1. raw visibility data — 783TB
2. calibration data — not significant
3. quick-look continuum images — 279TB
4. single-epoch continuum images — 314TB
5. single-epoch image cubes — 2918TB
6. single-epoch basic object catalogs — not significant
7. cumulative ”static sky” continuum images — 230TB
8. cumulative ”static sky” image cubes — 76874TB
9. cumulative ”static sky” basic object catalogs — not significant
10. **Total:** continuum images — 823TB; image cubes — 80PB

These volumes have been calculated assuming the angular and spectral resolutions defined in § 7, for the data products listed in § 4.1. We consider for the above calculation that the storage needed for the calibration data and catalogs to not be significant compared to the imaging and visibility data, and should be less than 10TB total.

For reference, the current JVLA archive has capacity for around 400TB. The normal expansion of this capacity, if driven by archiving of visibility data only, would be expected to be at the 100–200TB/year rate, and thus we would expect around 1.6PB of storage available in 2020. This does not include storage for images from future general image pipelines. We conjecture that support for storage of around 10PB by the end of VLASS is just within the practical envelope for funding within the NRAO budget, and even that assumes a extremely strong recommendation by the VLASS science review for this level of data support. Beyond this, partnerships with the community and other agencies will be necessary (see § 4.3).

The VLASS continuum image volume is within the range of expected storage capability. However, the 80PB of image cubes exceeds the currently extrapolated capacity of the JVLA archive by a factor of 50, and the wildly projected upper envelope by a factor of 8. Therefore, barring outside support through a EDS partnership, severe compression of the final cubes (in either angular or spectral resolution) is required.

Risk: Medium to High Details of the overall NRAO archiving plan as they extend into the future are yet to be developed fully and resourced. The VLASS archive should fit in with this overall plan, and perhaps be used as a testbed to explore options. The risk is mainly due to this uncertainty in overall NRAO data management strategy at this time. As noted above, the clear high (resource) risk issue is the large volume taken by the full-resolution image cubes. Fallback would be to impose severe reductions through compression and reduction in resolution on what can be served in real-time, and/or POD as needed.

We now estimate the data volumes required for storage of the individual VLASS data and data products:

1. raw visibility data
2. calibration data
3. quick-look continuum images
4. single-epoch images and image cubes
5. single-epoch basic object catalogs
6. cumulative ”static sky” images and image cubes
7. cumulative ”static sky” basic object catalogs

9.1 Raw Visibility Data

The VLASS will take a total of 9067 hours of observing time. If the entire survey used the standard full 2GHz bandwidth S-band WIDAR configuration with 0.5 second visibility integration times, then the data rate is 24MB/s. This means the total visibility data volume for the VLASS is 783 TB.

In practice, different parts of the survey may end up using different correlator configurations. For example, the Tier 1 and Tier 2 passes at shallow depth (141 μ Jy per pass) may use shorter integration times, increasing their data rates above 25MB/s. Likewise, the Tier 3 observations if not using OTFM could use longer integration times and thus lower data rates. Thus our 783 TB is an upper limit to the VLASS data volume.

Note that if it is necessary for us to reduce data rates and volumes further, we could employ frequency averaging in the correlator backend or archive. This has potential undesirable effects due to delay losses and loss of instantaneous field of view which could complicate processing, as well as limiting the usefulness of the data products for Rotation Measure analyses, and we do not recommend taking this approach unless truly necessary.

9.2 Calibration Data

The total primary and secondary calibration data for VLASS must come from the assumed 25% overhead time not on target, and in practice will be only around 10% of the total data volume. Furthermore, the calibration and flagging tables derived from this data will be signficantly smaller than this (being per antenna rather than per correlation), comprising much less than 1% of the total data volume. Thus we conclude that calibration data is an insignificant contributor to archive volume.

As noted earlier, the archive will need the facility to store, retrieve, and apply this calibration data to raw visibilities when requested by users and further processing pipelines.

9.3 Quick-Look Continuum Images

These images are produced for every scheduling block for purposes of QA and transient object detection. We assume for all image size calculations that the images will be pixellated at a sampling level 0.4 of the (robust weighted) resolution at the highest frequency of the band (4 GHz), rounded to a convenient value. For A-configuration, this is 0.2″ (resolution 0.49″), giving 324Mpix per square degree. For B-configuration, this is 0.6″ (resolution 1.58″), or 36Mpix per square degree. In practice we will refine this based on imaging performance and we may be able to get by with less oversampling. In addition, individual sub-images will likely need some amount of extra padding to accomodate odd shapes. Overall we should treat these estimates as reasonably conservative, uncertain at the around the 25% level.

For QL production images, we assume 8 BDP images (I, Ialpha, Q, U, V) plus uncertainties. See § refsec:qlimages for a description of these products.

The rates and volumes of data production depend upon the specific Tier:

9.3.1 Tier 1

The approximately 34000 square degress will be mapped in B-configuration in two epochs, with each epoch broken into a number of passes by sub-area. An all-sky image comprises 1.2Tpix, with size 5TB assuming single precision storage. In each epoch of the VLASS, the 8 QL images will take 40TB in two cycles. The Tier 1 QL total is 80TB for the images over the course of the survey.

9.3.2 Tier 2 — Extragalactic

The 10000 square degrees in B-configuration will be mapped in four epochs (see Scenario I in § 5.4). There are 360Gpix in this area, or each image is 1.44TB. Assuming we make the 8 canonical images for each QL epoch, we require 11.5TB per epoch (each epoch is single-pass over a given area). For 4 epochs the total storage required is 46TB.

9.3.3 Tier 2 — Galactic

This has 3160 square degrees in A-configuration for 1024Gpix (4TB) per image. The 8 BDP QL images require 32TB per epoch. There are four epochs, and at reasonable scan rates there can be no more than 2 passes per epoch. If we assume those passes are done to improve uv-coverage rather than provide an extra sub-epoch, then the four total epochs take 128TB total.

9.3.4 Tier 3 — COSMOS

The 2 square degrees in A-configuration has 648Mpix (2.6GB) per image. The 8 images in a given QL pass will take 20.8GB. There are 20 observing blocks per epoch, so each epoch will need 416GB. The total for the two epochs of COSMOS is a modest 832GB.

9.3.5 Tier 3 — ECDFS

The 4.5 square degrees in A-configuration takes 1.46Gpix (5.83GB). The 8 QL images per pass take 46.6GB. There are 98 passes per epoch (Scenario I in § 5.4) for 4.57TB per epoch. The total for 4 epochs is 18.3TB.

9.3.6 Tier 3 — Elias-N1

The 3.5 square degrees in A-configuration takes 1.13Gpix (4.54GB) per image. The 8 QL images per pass take 36.3GB. There are 39 passes per epoch (Scenario I in § 5.4) for 1.42TB per epoch. The total for 4 epochs is 5.7TB.

9.4 Single-epoch images and cubes

After each epoch, we will produce refined images and cubes. For these we assume there will be 10 wide-band continuum images (I, Ialpha, Q, U, V, plus uncertainty maps). As described in § 4.1.4, the image cubes can potentially have 1024 channels at full resolution (2MHz per channel). These could be very large for Tiers 1 and 2 and thus we only store these for Tier 3. In this case we expect to store only 896 channel cubes for IQUV plus the uncertainty maps, for 8 image cubes. We do plan to store coarse cubes (in the 14 viable spectral windows, each 128MHz wide) for all Tiers. These would be be for the 8 standard images (I, Q, U, V, plus uncertainty maps).

9.4.1 Tier 1

An all-sky image in B-configuration comprises 1.2Tpix, with size 4.8TB assuming single precision storage. In each epoch of the VLASS, the 10 continuum images will take 48TB. The 8 coarse (14 plane) image cubes will take 67.2TB per epoch. The Tier 1 total is 96TB for the continuum images and 134TB for the coarse resolution cubes.

9.4.2 Tier 2 — Extragalactic

There are 360Gpix in this area, or each image is 1.44TB. Assuming we make the 10 MFS continuum images, we require 14.4TB per epoch. The 8 coarse image cubes take 161.3TB per epoch. For 4 epochs the total storage required is 58TB for continuum images and 645TB for the cubes.

9.4.3 Tier 2 — Galactic

This has 1024Gpix (4TB) per image. The 10 BDP per-epoch continuum images require 40TB per epoch. The 8 coarse resolution cubes take 448TB per epoch. The four total epochs take 160TB total for the continuum images, and 1.8PB total for the cubes.

9.4.4 Tier 3 — COSMOS

The 2 square degrees in A-configuration has 648Mpix (2.6GB) per image. The 10 MFS continuum images in a given epoch will take 26GB. The 8 coarse (14 plane) cubes will take 291GB. The 8 full-resolution (896 channel) cubes will take 18.6TB. Note that we will also re-process the archival 2012-14 COSMOS data, which was taken in 3 epochs (two in A, one in C). The total for the five epochs of COSMOS is 130GB for the continuum images, 1.5TB for the coarse cubes, and 93TB for the full-resolution cubes.

9.4.5 Tier 3 — ECDFS

The 4.5 square degrees in A-configuration takes 1.46Gpix (5.83GB). The 10 continuum images per epoch take 58.3GB. The 8 coarse resolution cubes take 653GB. The 8 full-resolution cubes take 41.8TB. The total for 4 epochs is 233GB for continuum images, 2.6TB for the coarse cubes, and 167TB for full-resolution cubes.

9.4.6 Tier 3 — Elias-N1

The 3.5 square degrees in A-configuration takes 1.13Gpix (4.54GB) per image. The 10 continuum images per epoch take 45.4GB. The 8 image cubes take 508.5GB at coarse resolution and 32.5TB at full resolution. The total for 4 epochs is 140GB for continuum images, 2TB for coarse resolution cubes, and 130TB for full-resolution cubes.

9.5 Single-epoch basic object catalogs

The data volume for catalogs will not be significant. Flat catalog files in simple formats can easily be served from the archive. If served by a Relational Database (RDB), the number of entries may be large and may need consideration in implementation.

9.6 Cumulative ”static sky” images and image cubes

After each epoch after the first, we will produce refined images and cubes summed over the preceding decades, eventually arriving at the ”final static-sky image”. For these we assume there will be 12 wide-band continuum images (I, Ialpha, Icurv, Q, U, V, plus uncertainty maps), plus 8 coarse resolution (14 planes, 128MHz resolution) image cubes (I, Q, U, V, plus uncertainty maps). The full-resolution image cubes have 896 channels (2MHz per channel). We need to archive the coarse and full-resolution cubes only for the latest cumulative release.

9.6.1 Tier 1

An all-sky image comprises 1.2Tpix, with size 5TB assuming single precision storage. There are only two epochs, so there is a single final release set of cumulative images. The 12 continuum images will take 60TB. The 8 coarse resolution image cubes will take 560TB. The final full-resolution cubes will take 36PB.

9.6.2 Tier 2 — Extragalactic

There are 360Gpix in this area, or each image is 1.44TB. Assuming we make the 12 continuum images for each release, we require 17.3TB each. For 4 epochs there are 3 releases and the total storage required is 25.8TB for continuum images. A release for the coarse cubes takes 161TB, and 11PB for the full-resolution cubes.

9.6.3 Tier 2 — Galactic

This has 1024Gpix (4TB) per image. The 12 BDP per-epoch continuum images require 48TB per release. For 4 epochs there are 3 releases, adding 144TB total for the continuum images. A release for the coarse cubes takes 448TB, and 29PB for the full-resolution cubes.

9.6.4 Tier 3 — COSMOS

The 2 square degrees in A-configuration has 648Mpix (2.6GB) per image. There are two epochs so there is a single final release. The 12 continuum images will take 31.2GB. The 8 coarse cubes will take 291GB, and the full-resolution cubes 19TB.

9.6.5 Tier 3 — ECDFS

The 4.5 square degrees in A-configuration takes 1.46Gpix (5.83GB). The 12 continuum images per epoch take 70GB. The total 3 releases (4 epochs) is 210GB for continuum images. The final 8 coarse resolution cubes take 653GB, and the full-resolution cubes take 42TB.

9.6.6 Tier 3 — Elias-N1

The 3.5 square degrees in A-configuration takes 1.13Gpix (4.54GB) per image. The 12 continuum images per epoch take 54.5GB. The total for 3 releases (4 epochs) is 163GB for continuum images. The final 8 coarse resolution cubes take 653GB, and the full-resolution cubes take 42TB.

9.7 Cumulative ”static sky” basic object catalogs

The data volume for catalogs will not be significant. Flat catalog files in simple formats can easily be served from the archive. If served by a RDB, the number of entries may be large and may need consideration in implementation.

10 Test and Development Plan

There are a number of issues related to the VLASS that must be addressed before the survey can be observed on the telescope. We feel that none of these are ”show-stoppers” that are likely to prevent the survey from being carried out at all, and most have obvious work-arounds. Fundamentally, these are schedule and resource risks rather than functionality risks, in that it will take longer and will be more costly in computing and human resources to process the survey, impacting the data product delivery schedule. However, they do need to be addressed, and we propose a VLASS Test and Development Program leading up to and through the survey start. For example, short test observations or larger pilot observations are indicated in some areas, while analysis of archival data from previously observed projects such as Stripe-82 13B-370 will serve in others. These will require signifcant astronomer resources to carry out, and thus we are unable to fully execute this program before submission of the VLASS proposal — approval for the observation of VLASS would be necessary before allocating the resources to carry out this test program. This is particularly true for the issues in Tier 3 imaging. There would be a final critical design review before survey observations commence, and by that point we will have dealt with the high and medium risk issues sufficiently to proceed.

The areas we have initially identified for testing, exploration, and development include:

- General Flagging, Calibration, and Imaging issues
  - Tests of RFI occurence, RFI flagging efficacy, bandwidth losses to RFI over VLASS sky coverage area
  - Tests of performance of wide-band wide-field continuum imaging, assessment of achieved noise level, angular resolution, and ability to recover source spectral indices
  - Assessment of reliability of wide-field polarized beam maps (2–4GHz) and need for further measurements
  - Tests of general wide-band wide-field polarimetry, including application of beam maps to data
  - Determination of optimal vs. practical options for image pixel and spectral resolution and compression options (critical for archival storage planning)
  - Assessment of effects of source variability (intensity and polarization) on deeper static-sky imaging performance
- OTFM tests and development (primarily for Tiers 1 and 2)
  - Tests of imaging errors from OTF scanning (continuum intensity, spectral index, and polarimetry)
  - Tests of fast integration fast scan system and imaging performance (for depths shallower than 100 μ Jy rms)
  - Exploration of need for multiple passes in hour angle for uv-coverage improvements to imaging quality

  - Tests of Galactic Plane and Bulge region imaging (complicated source structures, crowded fields)
  - Development of improved algorithms required (if any) for OTFM imaging
  - Development of plan for small test or pilot observations for Tiers 1 and 2
  - Assessment of source finding algorithm performance for Tier 1 and 2 images, and identification of needed improvements

- Deep Mosaic Imaging tests (Tier 3)
  - Assessment of imaging performance for deep mosaicked fields (continuum intensity and polarimetry) based on archival data
  - Development of optimal scheduling for deep observations (e.g. scanning for best uv-coverage)
  - Development of improved algorithms required (if any) for deep imaging and polarimetry
  - Assessment of source finding algorithm performance for deep field images, and identification of needed improvements

- General Logistical Tests
  - Determination of observational overheads (slew, setup, calibration)
  - More detailed determination of actual increase in integration times needed for low-elevation observations than presented in this document
  - Exploration and testing of calibration issues for VLASS (e.g. sharing of calibration between blocks, cadence)
  - Exploration of optimal and practical scheduling options (fixed LST vs. dynamic)
  - Exploration and development of optimal/practical compression options for VLASS
  - Assessment and improvement of initial calibrator source lists for VLASS

- Calibration and Imaging Pipeline development and testing
  - Development and testing of basic calibration pipeline (including polarization calibration)
  - Development and testing of Quick-Look calibration pipeline
  - Development and testing of Quick-Look imaging pipeline
  - Development and testing of imaging pipelines for Tiers 1 and 2
  - Development and testing of imaging pipelines for Tier 3 imaging
  - Development and testing of source finding software for basic catalogs

Many of these are currently being carried out by VLA staff and resident observers as part of normal development and science support, and by the user community in their research activities. However, new resources (additional staff, post-doc, or student time, computing, and telescope test time) will be required to fully implement this plan before the start of the VLASS.

We also include here the development and testing of the pipelines for VLASS. These will be based on the general VLA pipelines that are now being deployed and used, tuned for VLASS specific cases. Testing on suitable projects (current and archival) will be an important aspect of pipeline testing and development.

We now describe more details pertinent to selected key areas:

10.1 General Flagging, Calibration, and Imaging Tests

The main tests to be performed are related to RFI, calibration, and wide-field polarimetric imaging performance of the JVLA. These are issues that are important for VLA observers in general, not just the VLASS. The VLASS provides us an opportunity for directed testing of these issues, with the results to be made available to the user community (e.g. through VLA memos and improved software tools).

Also to be studied in this area are the trade-offs that can be made in the image and image cube sizes (angular and spectral resolution) in order to reduce archive and distribution costs but still maximize the science capabilities of the survey. Also under investigation are the utility and limitations of more general lossless and lossy compression algorithms on the images.

We highlight a few of the key challenges below.

10.1.1 RFI, Bandwidth, and Sensitivity

As described in § 2.1, the main contribution to the sensitivity of the JVLA for continuum observations is the effective useable bandwidth that one must enter into the radiometer equation or the Exposure Calculator Tool. Past experience has shown that good (automated) imaging can be carried out with fairly brutal flagging of affected sub-bands, leaving around 1350 MHz of effective bandwith. On the other hand, past careful reductions have shown that 1500 MHz should be acheivable, and this is what we recommend to users. As this difference is equivalent to 11% in observing time or 5.4% in sensitivity, this issue constitutes a medium risk for the VLASS and will required significant testing as well as the development of automatic RFI heuristics (and possibly new algorithms).

Risk: Medium. Data can easily be obtained, but requires careful testing to establish the performance at all Declinations. Mitigation could involve loss of data and sensitivity, or development of more complex heuristics or algorithms.

10.1.2 Wide-band Wide-field Continuum Imaging

The substantial increase in the continuum survey speed for the JVLA comes from the greatly increased instantaneous bandwith provided by the upgrade. However, in order to produce a continuum image from this bandwidth, the system response and the source spectrum must be folded in to the calculation. This includes effects such as:

- the system response over the full bandwidth (e.g. the complex bandpass and system noise spectrum)
- the antenna primary beam responses (in the R and L polarizations) over the field-of-view (out into the near sidelobe)
- the intrinsic source spectrum over this band (intensity, spectral index, spectral curvature, etc.)
- the distribution of visibilities (and weights) in the uv-plane after gridding

These effects will all contribute to the final continuum image rms, resolution (”synthesized beam width”, sidelobe level and structure, response to extended emission, and reconstruction of the source spectrum (through the multi-Taylor-Term imaging).

The default way of dealing with the source spectrum over the bandwidth is to break the band into smaller coarse channels, such as imaging each 128 MHz spectral window separately as a sub-band. In this approach, the spectral index can be fitted across these sub-bands, and incorporated into the model. This can be done iteratively during imaging, until no more emission is detectable above the noise in each sub-band. The drawback is that deconvolution is limited to what is detectable in a sub-band rather than the combinded full-band image.

In principle, the technique of Multi-Frequency Synthesis (MFS) can be used to image the Taylor expansion terms (TT) with frequency over the full band. This is implemented in CASA as part of the `clean` task, with the performance as described in Rau et al. (2014). However, very recent tests using simulated point sources and the L-band CHILES data (Gim et al. private communication) may indicate at low signal-to-noise ratios (SNR) there may be a bias in the recovered spectral indices at $SNR < 20$. In this event, we would have to fall back to using the sub-band imaging approach described above.

Risk: Medium The tests should be straightforward, and software or algorithm flaws are identified CASA will need to apply bug fixes or re-engineering in any event. The fall-back to sub-band imaging may incur penalties in the achievable imaging rms.

10.1.3 Polarimetry

A key aspect of the VLASS as compared to previous surveys is the capability for wide-band polarimetry at sub-mJy flux density levels. This science case was highlighted in the VLASS White Paper "A Wideband Polarization Survey of the Sky at 2–4 GHz" Mao et al. (2014).

The use of the Jansky VLA for deep wide-band polarimetric is being pioneered through several RSRO programs. For example, graduate student Preshanth Jaganathan is carrying out C-band observations in Elias-N1 that will inform our development of VLASS processing. The CHILES-ConPol L-band procesing is also ongoing (C. Hales, NRAO). Of more direct interest are the COSMOS S-band observations (PI: Smolcic). These programs have identified issues and paths forward, in particular the need for high-quality full-Stokes primary beam measurements and models. Observations obtaining these and the processing are underway.

Risk: Medium to High. Given the unique science that polarimetry brings to the VLASS makes this a high science priority issue. That this JVLA capability is only recently being fully explored, makes this medium to high risk at the current time. Mitigation of this risk over the next year is a top priority test and development target.

10.1.4 Transients and Variability

Exploring the time domain is also a key science frontier for the VLASS. We have designed the survey in order to survey the sky in several epochs with sufficient cadence to identify candidate varying sources. The requirements on, and pitfalls of, a transient survey are described in Frail et al. (2012).

Previous and current JVLA programs (12A-371, 12B-158, 13B-370) are employing this capability and provide us a set of on-sky test cases. We will work with these groups to carry out the required tests for the VLASS.

Risk: Low Existing pilot programs are underway. No obvious show-stoppers. Main need is modest personnel time to investigate and document.

10.2 Tier 1 and 2 OTFM Testing

The VLASS science case in the wide and all-sky tiers is built around the efficiency of OTF mapping. The overhead time during OTF mapping is driven by time to slew to calibrators, which scales with time on sky. Overhead for pointed mapping scales with the number of points visited on the sky (beams times visits), so shallow and/or repeated visits to the same part of the sky would be very costly. However, OTF imaging has until recently been considered an experimental observing mode, so more work is needed to confirm in detail that the VLASS science goals are achievable in this mode. Here we describe the open questions to achieving the full VLASS science case (particularly transients) via OTF mapping and a plan to resolve these questions.

As detailed in the Guide to Observing with the VLA, survey speed is equal to the beam area divided by the integration time needed to reach a given sensitivity:

$$SS(\mathrm{deg}^2\mathrm{hr}^{-1}) = 0.5665\,\theta_{pb,ref}^2/t_{int} \tag{3}$$

where $\theta_{pb,ref} = 15'$ is the primary beam FWHM at the band center reference of 3 GHz, and t_{int} is the integration time reported by the VLA Exposure Calculator Tool for the required image rms. In OTF mode, the survey area is covered by a series of stripes ("rows") spaced by θ_{row}. For nearly uniform sensitivity coverage over the band, we adopt a value

$$\theta_{row} = \theta_{pb,min}/\sqrt{2} = 0.53\,\theta_{pb,ref} \tag{4}$$

where $\theta_{pb,min} = 0.75\,\theta_{pb,ref} = 11.25'$ is the primary beam FWHM at the upper frequency of 4 GHz. This striping parameter translates the survey speed to an antenna slewing speed relative to the sky:

$$\begin{aligned}
\dot{\theta} &= SS/\theta_{row} && (5)\\
&= 1.07\,\theta_{pb,ref}/t_{int} && (6)\\
&= 1.43\,\theta_{pb,min}/t_{int} && (7)\\
&= 2.08\,\mathrm{arcmin\,s}^{-1}\,(\sigma_I/\sigma_{I0})^2 && (8)
\end{aligned}$$

for our nominal survey speed of $SS = 16.55$ arcmin^2/sec for image rms of $\sigma_{I0} = 100\mu$Jy/beam.

In an ideal case, OTF enables any survey to be conducted as N-epochs, each with depth $\sqrt{N}$ of the final depth. The VLASS science case assumes these can be imaged individually and jointly to produce both deep maps and information about the transient radio sky. However, a critical limitation is that the correlator dump time must be fast enough to temporally resolve the amplitude envelopes of sources as they move through the primary beam pattern. The dump time must sample sources roughly faster than 10 times its motion through the primary beam:

$$t_{\mathrm{dump}} * \dot{\theta} < 0.1\,\theta_{pb,min}. \tag{9}$$

Plugging in $\dot{\theta}$, relates t_{dump} to overall integration time and the combined sensitivity:

$$t_{\mathrm{dump}} < 0.093\,t_{int}. \tag{10}$$

The minimum dump time (and thus number of possible epochs) is limited by the output data rate, which should not exceed 25 MB/s. The VLA produces data at a rate R_{data} of:

$$R_{data} = 45 * (n_{spw} * n_{ch} * n_{pol}/16384)/t_{\mathrm{dump}}. \tag{11}$$

Conservatively assuming that all 16 subbands are useful (2 GHz bandwidth), each with 64 channels and 4 polarization products, the VLASS will produce 11.25 MB per integration. For a data rate limit of 25 MB/s, the shortest full-VLASS dump time is 0.45s and the shortest integration time per point on the sky is 4.8s for the nominal VLASS configuration bringing back the full 2048 MHz bandwidth ($n_{ch} = 1024$). This corresponds to a minimal sensitivity of 126μ/beam image rms per pass. Note that this minimal sensitivity does not depend on field of view, but does assume that OTF stripes are spaced optimally as described above, that the full available correlator bandwidth is used, and that the data rate limit must be obeyed at all times.

The Tier 1 ALL-SKY of the VLASS proposes a final sensitivity of 100μJy/beam in a single pass, while Tier 2 proposes a final depth of 50μJy/beam with a depth of 100μJy/beam in each of four passes. This is acheivable within the constraints described above. However, it would be highly desirable to carry out the ALL-SKY in two passes each with depth of 141μJy/beam (either two epochs or two different hour angles for improved uv-coverage). Likewise, it may be desirable to observe Tier 2 with two hour-angle passes in each of the four epochs.

Since the ALL-SKY tier in particular would be vastly more transient-capable with at least 2 epochs, it is worth considering how to accommodate a shallower survey depth per pass. The options are:

1. lower the data rate by not storing the full bandwidth, dropping whole spectral windows (out of the 16 total) to redude the rate,
2. spread the rows of the OTF raster further apart (e.g. by 4/3 to be optimally spaced at mid-band rather than the top of the band),
3. and/or by trading off having higher data rates during the Tier 1 and/or Tier 2 observing against lower rates during the deeper Tier 3.

Option (3) is a policy decision, we will assume the policy stands as enforcing data rate limits at all times. Option (2) would trade off non-uniformity in the final mosaic sensitivity versus the data rate. This will require simulation and testing to verify, but one might expect that uniformity of the wideband MFS images would still have good uniformity for modest spacing increases (10%?). Finally, option (1) is viable given that a number of spectral windows are largely contaminated by RFI and thus dropping them entirely is reasonable. However, dropping a spectral window precludes any recovery of data via flagging (RFLAG or TFCROP). There are two S-band spectral windows almost entirely lost to RFI, and a further two with greatly diminshed sensitivity over much of their span due to band roll-off. Dropping 4 windows out of 16 (still retaining 1536 MHz bandwidth) will allow a 33% increase in integration time or a 15% higher rms sensitivity limit per pass. Thus this alone would enable a per-pass depth of 145μJy/beam, which is sufficent to make two passes for ALL-SKY and each Tier 2 epoch. This option will require some testing to make sure that the correlator back-end can deliver data for long periods at the short dump times and that the lost data does not reduce the image sensitivity below the desired values.

With this sketch of how a multi-epoch OTF survey could work, we outline the open questions for this mode and our plan to address them:

1. What imaging errors (and what scale) are introduced by fast OTF slewing?
    - We have test data taken at a range of slew rates that will define the magnitude and scaling of imaging artifacts for OTF mode.
    - This test will give us a maximum safe slew rate for a given required image sensitivity. Given this limit, we can define the maximum slew rate and whether two epochs are possible for the all-sky tier.
2. Does the sensitivity of co-added epochs scale ideally down to 50 microJy?
    - Science (PI: G. Hallinan) and test OTF data in S-band are in hand to study whether shallow co-added epochs can reach the theoretical sensitivity at a level of 50 microJy.
    - This test will benefit from ongoing development of joint deconvolution image cleaning in CASA.
3. Can the correlator observe with 16 spectral windows and a dump time of 0.34 s?
    - A correlator test observation would quickly answer whether this is possible.
4. What spectral windows will be significantly hit by RFI for an all-sky survey?

    - Search of archival data for S-band RFI.
    - If a spectral window is unusable for a large fraction of the sky, we will exclude it and recalculate the minimal single-epoch sensitivity.

5. What polarization bias/uncertainty is introduced by OTF mapping?
    - Need new observing of a polarization calibrator in OTF and pointed mapping modes.
    - Compare polarization bias as a function of frequency and pointing location. Polarimetric mapping in both modes requires some algorithm development to accommodate large fractional bandwidths.
    - More observing needed to generate calibration products for all polarimetric calibration in general.

6. What computing resources are needed for Tier 1 and 2 image processing?

There are a number of more high-level issues that also need to be addressed in order to plan and resource the execution of Tiers 1 and 2:

1. What, if any, are the required algorithm developments (and development and testing resources) needed to carry out Tier 1 and 2 imaging and (self) calibration?
2. What are the additional issues for the imaging of the Galactic Plane and Bulge regions in Tier 2?
3. What computing resources are necessary to image Tiers 1 and 2?
4. What are the automated or semi-automated image pipeline scripts that will be necessary for Tiers 1 and 2, and what pipeline infrastructure developments will be necessary beyond minor modifications of the scripts and process used in 13B-370?
5. What personnel resources are necessary to carry out the imaging and quality control for Tiers 1 and 2?
6. What, if any, small test or larger pilot observations are needed to prepare the plan for Tiers 1 and 2?

10.3 Tier 3 Deep Imaging Tests

The Jansky VLA has successfully been used to take data, which has been processed using AIPS and CASA (sometimes in combination) make deep ($\sim 1\,\mu$ Jy rms) images from wideband continuum data at S-band. However, deep imaging at this level is not routine and is heavily manual in nature. Most of the past experience has been in imaging single pointings (Owen et al. A2256 and GOODS-N, Van Gorkom et al. CHILES COSMOS pointing at L-band). The most relevant current mosaic is the 12B-158 Smolcic et al. COSMOS mosaic that is the basis for the Tier 3 COSMOS field. This is currently being processed by the Smolcic group, and lessons learned in this processing will be critical to developing the plan and pipeline for Tier 3 processing.

The VLASS Deep Fields will present a challenge to our ability to reliably and efficiently pipeline image deep wideband polarimetric mosaics. The key high-level issues that we foresee include:

1. Is the CASA software able to fully image the Tier 3 data in wide-band continuum and polarization?

2. Is the VLA instrument (e.g. primary beam) sufficiently characterized to allow imaging and calibration of the wide-band wide-field polarimetric mosaics at the depth and dynamic range required for Tier 3?
3. What, if any, are the required algorithm developments (and development and testing resources) needed to carry out Tier 3 imaging and (self) calibration?
4. What computing resources are necessary to image Tier 3?
5. What level of automated or semi-automated pipelining will be possible for Tier 3? What are the implications of having to fall back to extensive manual image processing?
6. What personnel resources are necessary to carry out the imaging and quality control for Tier 3?
7. What, if any, small test or larger pilot observations are needed to prepare the plan for Tiers 3?

These issues are similar to those for Tiers 1 and 2, but are included here because the answers are likely to be different in detail.

10.4 General Logistical Tests

The main logistical issues to be tested and verified are:

1. What are the observational overheads (slew, setup, calibration) for the VLASS? Initial estimates based on past experience and construction of example schedules were described in § 2.3. Verification of these assumptions on the basis of real VLASS schedules is required.
   - Overheads for Tier 1 (All-Sky) using OTF scanning and fixed or long-block scheduling.
   - Overheads for Tier 2 (Wide Extragalactic and Galactic) using OTF scanning with fixed or long-block scheduling.
   - Overheads for Tier 3 (Deep Fields) using dynamically scheduled blocks and:
     - OTF scanning
     - Hex-pattern pointed mosaics

   Note: the ability to share calibration between dynamically scheduled short blocks in Tier 3 will factor into the overhead tests.
2. What increase in integration times are needed to reach uniform depth for low-declination fields?
   - Sensitivity loss for Tier 2 Galactic in the:
     - Galactic Bulge region
     - Galactic Plane region
   - Sensitivity loss for Tier 3 ECDFS field

   These would be carried out through test observations in the regions under study to see the loss of sensitivity in imaging compared to equivalent observations at higher elevation. In addition to these direct imaging rms tests, we will use the ”Switched Power” data in the VLA dataset to track the system temperature variations.

11 Project Schedule and Resourcing

11.1 Schedule

The key high-level **milestones** and notional dates, as well as *opportunities for testing*, following the March 2015 VLASS Review and a subsequent decision by the NRAO Director to carry out the VLASS, and leading up to the start of VLASS observing, are:

- **2015 March** VLASS Project starts. Establishment of Project Office and initial staffing.
- *2015 Feb 6 – 2015 May 11* 2015A B-configuration. Opportunity for on-sky tests of VLASS B-configuration observing, as well as calibrator searches.
- *2015 May 15 – 2015 Jun 1* 2015A BnA-configuration. Opportunity for on-sky tests of VLASS BnA hybrid configuration observing.
- *2015 Jun 12 – 2015 Sep 21* 2015A A-configuration. Opportunity for on-sky tests of VLASS A-configuration observing, as well as calibrator searches.
- **2015 June 15** Face-to-face VLASS Conceptual Design review. Mostly a follow-up on issues identified in review of the preliminary TIP (this document).
- **2015 October 15** VLASS Preliminary Design review.
- **2016 April 15** Final VLASS Critical Design review. Positive outcome of review will be a ”go” for commencing observing, possibly with modifications of plans and delay of start.
- *2016 Feb 5 – 2016 Apr 25* 2016A C-configuration. Opportunity for final on-sky tests and preparatory calibrator searches.
- *2016 May 27* Start of 2016A observing in B configuration, and earliest possible start for VLASS observing assuming no modifications to VLA configuration cycles subsequent to this.

11.2 Resource Requirements

The telescope time, computing resources required for data processing, and archive storage, have been described in detail above. Here we address the effort required to execute the Test and Development Plan, and to deliver the VLASS and its basic data products. These are FTE estimates based on local experience and expertise, and include an overhead to account for the time NRAO scientific staff are expected to spend on their self-directed research.

Test and Development Plan, as described in § 10 (March 2015 through April 2016):

- General flagging, calibration, and imaging issues
  - 0.4 FTE-yr (risk: low)
- OTFM tests and development
  - 0.3 FTE-yr for tests, unknown level of effort needed if improved OTFM imaging algorithms are found to be required (risk: medium)
- Deep Mosaic Imaging tests (Tier 3)

  - 0.2 FTE-yr for scheduling and imaging assessments, unknown level of effort needed if improved imaging algorithms are required (risk: medium)

- General logistical tests
  - 0.3 FTE-yr (risk: low)
- Calibration and imaging pipeline development and testing
  - 0.5 FTE-yr (risk: low)

The total effort required for the Test and Development phase of the VLASS is 1.7 FTE-yr, plus imaging algorithm development (if the latter is found to be required).

VLASS operations and production of BLDs (May 2016 – Jan 2021):

- Management
  - 0.2 FTE/yr (risk: low)
- Scientific staff to oversee observing execution and pipeline products
  - 0.5 FTE/yr (risk: low)
- Data analysts to oversee pipeline product quality assurance
  - 2 FTE/yr (risk: low)
- Scientific staff to develop and oversee catalog production
  - 0.3 FTE/yr (risk: high)

The total effort required during VLASS operations is 3.0 FTEs/year.

Development of "processing on demand" capabilities:

- Requirements definition (scientific staff)
  - 0.2 FTE-yr (risk: medium)
- Software development and implementation
  - 1.5 FTE-yr (risk: medium)
- Testing (scientific staff)
  - 0.2 FTE-yr (risk: medium)

The total effort needed in order to implement POD is estimated to be 1.9 FTE-yr.

These numbers reflect the NRAO staff time required to carry out or supervise these activities. There are a number of areas that can be addressed and augmented through the participation of members of the astronomical community (SSG members, post-docs, graduate students, NRAO REU students in a few cases). We plan to advertise these opportunities and incorparate participants into the project where possible and practical. Note that use of community shared effort in these technical areas is unlikely to significantly reduce the required NRAO staff levels given above, as careful coordination and support will be required.

References

Frail, D. A., Kulkarni, S. R., Ofek, E. O., Bower, G. C., & Nakar, E. 2012, Astrophysical Journal, 747, 70

Hancock, P. J., Murphy, T., Gaensler, B. M., Hopkins, A., & Curran, J. R. 2012, Monthly Notices of the Royal Astronomical Society, 422, 1812

Mao, S. A., Banfield, J., Gaensler, B., et al. 2014, ArXiv e-prints, arXiv:1401.1875

Mooley, K. P., Frail, D. A., Ofek, E. O., et al. 2013, Astrophysical Journal, 768, 165

Myers, S. T., Jackson, N. J., Browne, I. W. A., et al. 2003, Monthly Notices of the Royal Astronomical Society, 341, 1

Rau, U., Bhatnagar, S., & Owen, F. N. 2014, ArXiv e-prints, arXiv:1403.5242

Winn, J. N., Patnaik, A. R., & Wrobel, J. M. 2003, VizieR Online Data Catalog, 214, 50083