1 Title: The VISTA Kilo-degree Infrared Galaxy survey (VIKING)

PI: W. Sutherland, Queen Mary, University of London (QMUL), U.K.
On behalf of the VIKING collaboration.

1.1 Abstract

The VIKING survey aims to survey two stripes of high galactic latitude sky totalling $\approx 1500\,\mathrm{deg}^2$, in five near-IR passbands (Z,Y,J,H,K_s) with typical on-source time $300-500$ sec per passband, thus substantially deeper (+1.5 mag) compared to the UKIDSS-LAS or VISTA-VHS. Combined with visible *ugri* bands from the VST KIDS survey of the same areas, this will yield a 9-band survey with a unique combination of depth and area.

Our science aims are wide-ranging, from photometric redshifts for cosmology, dark energy and weak lensing; quasars at $z > 7$; galaxy evolution at moderate redshift; galaxy morphology studies; galactic structure; and ultracool brown dwarfs.

In addition to the primary aim of complementarity with KIDS, our two selected stripes comprise the optimally-located regions of each Galactic cap from Chilean sites, and all of it overlaps with 2dFGRS and/or SDSS. There is substantial overlap with the GALEX Medium Survey, and with planned future surveys including South Pole Telescope, Dark Energy Survey, Pan-STARRS and AAOmega.

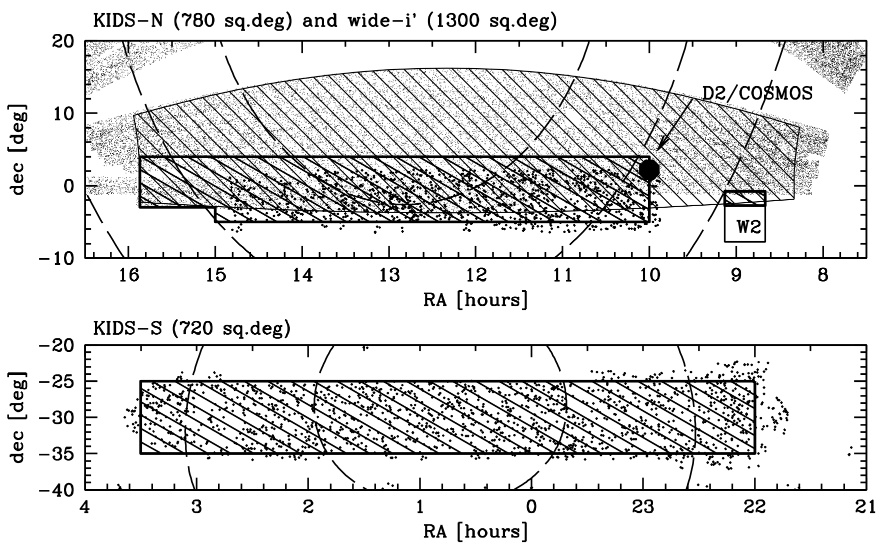

Figure 1: The sky coverage of the VIKING and KIDS surveys (thick line) in the North (top) and South (bottom) galactic caps. The CFHLS-W2 and COSMOS fields are labelled. A random subsample of SDSS (small dots) and 2dFGRS (large dots) redshifts are plotted.

2 Survey Observing Strategy

2.1 Scheduling requirements

Our choice of areas (two stripes , one each in SGP and NGP) is clear for several reasons: the two stripes give observability for some part of the survey at any sidereal time.

The stripes are essentially the "best" regions of high-latitude sky in the two galactic caps, since the SGP stripe is ideally located just South of the Paranal zenith (minimising wind and Moon interference), while the NGP stripe on the Equator is the best practical compromise between airmass, Galactic extinction, and 2dFGRS and SDSS coverage.

The approximate coordinate ranges are:

NGP: $10^h < \mathrm{RA} < 15^h$, $-5^\circ < \delta < +4^\circ$, plus $15^h < \mathrm{RA} < 15^h50$, $-3^\circ < \delta < +4^\circ$.

SGP: $22^h < \mathrm{RA} < 03^h30$, $-36^\circ \lesssim \delta \lesssim -26^\circ$. (The KIDS SGP stripe may be shifted Southwards by a small amount $\sim 1-2$ deg to maximise overlap with South Pole Telescope. Exact boundaries are still to be confirmed, but we assume a 1 deg shift for illustration).

We intend to locate our field centres along "tramlines" of constant Declination matching those of KIDS, using the default Camera position angle so the long (X) axis of the VISTA IRCAM field is East/West, and the short (Y) axis is North/South. Our Dec tramlines will be identical to KIDS at 59 arcmin pitch, which is just smaller than the 61 arcmin Y-width of one VISTA tile. Thus, each single RA stripe of VIKING will cover one KIDS stripe with a small excess, and after observing multiple stripes we will get useful 2 arcmin overlaps (of the full-depth VISTA tile) between neighbouring stripes.

The RA spacings must of course be different to KIDS due to the wider VISTA X-axis; we baseline a 1.46 deg pitch giving 1 arcmin overlaps between neighbouring VISTA tiles at the East/West edges. Our stripe dimensions are then 49×10 tiles (SGP) and 61×9 tiles (NGP), for a total of 1039 tiles. The unique area per VIKING tile is then $1.46\,\mathrm{deg} \times 0.983$ deg $= 1.435\,\mathrm{deg}^2$, hence a total area $1490\,\mathrm{deg}^2$.

Following the recommendation of the PSP, we aim to complete the full area over 5 years / 10 periods, rather than the 4 years in our September 2006 proposal submission. We note that our 2-year plan will contain the 'highest priority' subset of $600\,\mathrm{deg}^2$ with maximal overlap with other ongoing surveys: this comprises the KIDS+2dFGRS+SPT+DES mutual overlap area in the SGP at $-30^\circ < \delta < -36^\circ$, and the KIDS+2dFGRS+SDSS+GALEX-MIS mutual overlap in the NGP at $-2^\circ \lesssim \delta \lesssim 2^\circ$.

Some part of our area is observable at **any** sidereal time, so in the early stages of the survey time can be scheduled at any month. In the long term, we need time somewhat weighted towards Spring and Autumn seasons and less in Summer/Winter (since, if time were uniformly distributed across seasons, the ends of the stripes will finish earlier leaving underfilled patches in the middle).

Also, our seeing requirement of ≤ 1.0 arcsec is moderate, and there are no specific 'synoptic' aims in the current plan, thus overall our survey offers good flexibility with respect to VISTA scheduling.

There is clearly a compromise needed between the desire to minimise time-lags between observing the same field in different filters, (to minimise colour errors due to variable/moving objects) and the desire to observe redder filters nearer to Full Moon for observing efficiency. Our requested strategy, which we believe is an optimal compromise, is that each (filled) VISTA tile will be completed in **two visits of one observing block (OB) each** : the Z and Y observations will be put into one OB, the H and $\mathrm{K_s}$ put into a second OB, while our total J-band exposure time (400s) will be cut into two halves shared between the two OBs above.

Thus, each tile comprises two OBs, a "$\mathrm{ZYJ_1}$" OB requiring dark or grey Moon, and a "$\mathrm{J_2HK_s}$" OB which can use any Moon phase. These two OBs can be observed in either order. It is **desirable** that both OBs for one tile should occur within one month of each other, but this is **not** a requirement. We plan to use TOBOGAN to increase the score for tiles with one of their two OBs completed to assist in this.

Given the above strategy, objects showing significant variation in J-band between the two visits can be flagged, and cut if desired, so we can minimise contamination of rare-object candidates (colour outliers) by variable objects.

Since the Z-band observations for VIKING were originally part of the KIDS proposal and have been moved to VIKING at the request of the Public Surveys Panel, it is clearly important that VIKING Z-band does not lag significantly behind KIDS progress; therefore, we will increase the priority of "ZYJ_1" observing blocks for tiles which have already been observed by KIDS.

Since there are **no absolute timing** constraints, the relative timing constraint between two OBs for a tile is a soft constraint, and there is no coupling between timings for different tiles, our 2090 OBs essentially comprise 1045 independent pairs with differing priority scores, and these scheduling requirements appear straightforward to implement.

We note also that for proper motions of brighter objects, earlier-epoch data (although 1.5 - 2 mag shallower) will be available for the NGP stripe from SDSS z-band and ongoing UKIDSS data. For SGP proper motion data, a later-epoch pass on our SGP stripe (e.g. a 3rd visit with 200sec at J-band only) could be considered after the end of the nominal VIKING plan, taking $\sim$ 125 hrs, but this is left as a future option and not included in the current request.

Period	Time (h)	Mean RA (h)	Moon	Image FWHM (arcsec)	Transparency
P80	114	02, 11.5	dark/grey	≤ 1.0	thin
P80	106	02, 11.5	any	≤ 1.0	thin
P81	114	14.5, 23	dark/grey	≤ 1.0	thin
P81	106	14.5, 23	any	≤ 1.0	thin
P82	114	02, 11.5	dark/grey	≤ 1.0	thin
P82	106	02, 11.5	any	≤ 1.0	thin
P83	114	14.5, 23	dark/grey	≤ 1.0	thin
P83	106	14.5, 23	any	≤ 1.0	thin
P84	114	02, 11.5	dark/grey	≤ 1.0	thin
P84	106	02, 11.5	any	≤ 1.0	thin
P85	114	14.5, 23	dark/grey	≤ 1.0	thin
P85	106	14.5, 23	any	≤ 1.0	thin
P86	114	02, 11.5	dark/grey	≤ 1.0	thin
P86	106	02, 11.5	any	≤ 1.0	thin
P87	114	14.5, 23	dark/grey	≤ 1.0	thin
P87	106	14.5, 23	any	≤ 1.0	thin
P88	114	02, 11.5	dark/grey	≤ 1.0	thin
P88	106	02, 11.5	any	≤ 1.0	thin
P89	114	14.5, 23	dark/grey	≤ 1.0	thin
P89	106	14.5, 23	any	≤ 1.0	thin
Total dark/grey: 1140 h ; total any: 1060 h					
Grand total : **2200 h** , 10 periods					
Useful nights : 244 (assumed 10h night minus 1h calibration)					

Table 1: Distribution of observations over observing period and conditions. Note that each of our two strips straddles the semester boundaries, so we give two representative RA quartile points for each stripe above. Observing times **include** all estimated overheads associated with each OB (slewing, acquisition, jittering, detector readouts, etc), but **exclude** overheads for the VIRCAM Calibration Plan. Nighttime calibration overheads thus reduce the available hours per night pro-rata among all programmes.

2.2 Observing requirements

As outlined above, our strategy requires 2 visits to complete each tile, a "ZYJ_1" visit in dark or grey Moon phase, and a "J_2HK_s" visit at any Moon phase, in either order. The seeing requirement is ≤ 1.0 arcsec, and the sky transparency requirement is THIN. Each visit can complete one filled VISTA tile with a standard 6-pawprint pattern in the respective "2.5 bands" in approximately 1 hour including overheads (see Table 2 for more details).

Within a visit, all 6 pawprints will be observed first in one filter, then the 6 pawprints repeated in the second filter, then the third; this both minimises sky brightness changes within one filter and also reduces overheads, since a filter change will take longer than a telescope offset ~ 11 arcmin between pawprints.

Within each passband and pawprint, we will observe a 2-, 3- or 4-point jitter pattern (2 at J, 3 at H, 4 at Z,Y,K_s) with jitter steps of ≈ 20 arcsec, for optimal removal of sky background and bad pixels. Microstepping will not be used, since 2 VIRCAM pixels = 0.68 arcsec and our seeing requirement is significantly larger than this. A single exposure of 50 or 60 sec (according to passband) will be taken per jitter position, with values of NDIT and DIT dependent on passband, e.g. a single DIT for Z,Y and 5 or 6 10-sec DITs for H,K_s.

We note here that our strategy provides a good level of redundancy, since each single object will be imaged on 4, 6 or 8 detector locations per passband in one OB, enabling good outlier rejection. Likewise, within one OB, each single detector pixel will see 12, 18 or 24 disjoint sky positions in each single filter, enabling good background mapping by stacking and object rejection even if only one VIKING OB is observed on a given night.

We note that the J and H sky backgrounds (dominated by OH lines) are higher on average in the early part of each night, while the K_s background is dominated by thermal emission and has little correlation with time of night; also, K_s data can be taken up to ~ 20 mins from sunset/sunrise. Therefore, within a visit, the filters may be observed in arbitrary order, so for example the K_s filter may be observed first at evening twilight or last at morning twilight to optimise efficiency.

(Optionally, for some tiles, we may split our single J_2HK_s OB into two parts, with K_s becoming a third OB of approx 25 minute duration. This may be preferred operationally for flexible scheduling (minimising quantisation problems) and optimising use of morning twilight. This option will be left at the discretion of the Paranal operations staff, i.e. we will discuss it with them and implement it if they decide it is significantly helpful during operations.)

As noted above, we intend to prioritise so that areas with existing KIDS data, and/or South Pole Telescope coverage, and/or planned data from other surveys (DES, GALEX-MIS, AAOmega etc) are given highest priority, thus maximising area with optical-IR and spectroscopic coverage, plus radio and UV to the extent available.

2.3 The SDSS/UKIDSS overlap

We note that our NGP stripe overlaps with SDSS and the ongoing UKIDSS Large Area Survey. We emphasise that this is complementary rather than redundant, since the VIKING exposures are $8-12\times$ longer than UKIDSS: when combined with the $1.5\times$ higher QE of VISTA this will reach some 1.5 magnitudes deeper. Compared with SDSS $z-$band the improvement is even more dramatic, with $10\times$ the exposure time, $2.5\times$ the collecting area, $3\times$ the QE and superior seeing leading to a gain ~ 2.3 magnitudes.

The overlap with SDSS/UKIDSS is of course valuable for cross-calibration and proper motion data on the brighter objects, and this equatorial stripe is much superior to any possible 'new' NGP region e.g. $-20 \lesssim \delta \lesssim -10$ due to lower extinction, 2dF/SDSS redshift coverage for galaxy-galaxy lensing, and planned GALEX-MIS and AAOmega coverage, among others.

	Z	Y	J	H	Ks
Time & depth on sky in coadded Tiles					
Vega or AB mags?	Vega	Vega	Vega	Vega	Vega
Depth (mag) required	22.6	21.7	21.3	20.2	19.4
Sigma required	5.0	5.0	5.0	5.0	5.0
Assumptions					
In band sky brightness assumed - Vega mag/arcsec	18.4	17.2	16.0	14.0	13.0
Detector Integration Time (DIT) sec used in ETC	50	50	25	10	10
Time (sec) required per object assuming above values	500	400	400	300	480
Area required sq. deg	1476	1476	1476	1476	1476
Tiles required to cover area(s)	1029	1029	1029	1029	1029
effective useful sq deg/tile	1.43	1.43	1.43	1.43	1.43
Assign priorities to different areas?	yes	yes	yes	yes	yes
Parameters set					
DIT already assumed above	50	50	25	10	10
Ndit - # of Exposure coadds	1	1	2	5	6
Nexp # of Exposure loops	1	1	1	1	1
Nmicro- # of microstep positions (steps <3 arcsec)	1	1	1	1	1
Njitter - # of Jitter positions (steps odd # of 0.5 pixels < 30 arcsec)	5	4	2	3	4
Npaw - # of Pawprints in tile	6	6	6	6	6
Observe tile in same OB how many times?					
Number of filters in same OB? If >1 which other?	2.5	2.5	2.5(*)	2.5	2.5
Number of tile positions in same OB?	1	1	1	1	1
Resulting values					
Total Exposure sec/tile	1500	1200	600	900	1440
Total Elapsed sec/tile	1757	1410	725	1133	1770
Total Elapsed mins/tile	29.3	23.5	12.1	18.9	29.5
Observing efficiency %/tile	85.4	85.1	82.8	79.4	81.4
Time per object for s-to-n -single OB	500	400	200	300	480
Signal to noise (at depth required in row 3) - single OB					
Depth (to sigma in row 4) - single OB					
Multiple Tile Strategy					
# of Tiles per filter for S/N	1	1	2	1	1
Total Elapsed Hours per filter	502.2	403.0	414.5	323.8	505.9

Table 2: Observing strategy spreadsheet: summarising our required magnitude limits (all 5σ, Vega scale in 2 arcsec aperture, at 1.2 airmass; with ETC-derived exposure times, jitter and tiling strategy, DIT values and elapsed times per filter.

3 Survey data calibration needs

We anticipate that the standard VISTA calibration plan will be adequate for VIKING observations. We refer to the Calibration Plan, Ref. [01], for further details, but summarise the main steps here:

3.1 Instrumental signature removal

Ref. [01] specifies the basic instrument calibration frames (dark frames, reset frames, dome flats and linearity calibration, twilight flats) which will be available as part of the VISTA standard operating procedure, mainly from daytime and twilight procedures.

- Dark frames: one set for each typical DIT value expected to be taken approximately daily to weekly (as needed) during daytime.
- Dome flats: one set expected to be taken weekly during daytime, (primarily for detector health checks and logging of bad pixels).
- Linearity frames: these comprise sets of dome flats with stable illumination spanning a wide range of different exposure times. We anticipate a set of linearity frames will be taken every few months, and after any major maintenance on the IRACE detector controllers.
- Twilight flats will be taken in several filters nightly. (There is likely to be insufficient twilight time to obtain twilight flats in all science filters every night, but a cyclic ordering should get twilight flats in all filters every 2 nights, with priority given to those passbands most used recently) .
- "Touchstone" fields will be observed approximately two-hourly each night. A network of suitable non-crowded fields, either 2MASS touchstone fields or UKIRT faint standards, will be set up spaced at 2-hour RA intervals, both North and South of the Paranal zenith to minimise Azimuth slew overheads.

We note here that both the VISTA dome and VIRCAM dark filter have excellent light-tightness by design (in fact, VIRCAM can take dark frames in a normally-lighted lab), so daytime darks and dome flats should not have significant leaks. This will be checked during VISTA commissioning.

3.2 Astrometric calibration

The basic astrometric procedure is a 2-stage calibration: firstly, a radial distortion correction of the form

$$r_{\rm lin} = r + k_3 r^3 + k_5 r^5$$

is fitted, where r is image radius (in mm), k_3 and k_5 are distortion coefficients, $r_{\rm lin}$ is a linearised radius proportional to $\tan\theta$, where θ is angle from the pointing axis ; this form is an excellent fit to real distortion in axisymmetric systems.

The distortion parameters k_3, k_5 are calculated on a one-time basis based on fitting to a stack of a large number of frames of moderately high stellar density, including "mesostepped" frames with offsets ~ 0.5 detector width.

Then, for every data frame, a catalogue of bright unsaturated stars is produced, with pixel centroids, and matched to 2MASS using the approximate pointing information in the FITS headers as a start-point. The radial distortion correction is applied to give linearised coordinates x_l, y_l for each measured object, then a standard 6-parameter "plate constant" solution is performed, of the form

$$\xi = a x_l + b y_l + c \qquad \eta = d x_l + e y_l + f$$

where ξ, η are standard tangent-plane coordinates centred on the telescope pointing axis. This allows for pointing error, rotation, scale change and shear. The coefficients $a \ldots f$ are fitted with a robust fit to minimise observed–predicted residuals on a per-detector basis. The above astrometric solution will then be stored in the Multi-extension FITS headers in the ICRS system, using the `ZPN` notation to handle the distortion.

This procedure is very similar to that used currently for WFCAM data, which is demonstrated to give residuals over the whole field to less than 0.1 arcsec systematic and 0.1 arcsec random rms (with the latter limited by 2MASS random errors at its moderate SNR ~ 10).

While VISTA has a larger absolute distortion term, we anticipate that astrometric stability across the VISTA focal plane is likely to be at least as good or probably better than equivalent WFCAM results, for several reasons:

1. VIRCAM covers 3× the area of WFCAM per single pawprint, giving correspondingly more useful 2MASS stars per frame.
2. VIRCAM's detectors are firmly attached to a common CTE-matched mounting plate, rather than held in ZIF sockets.
3. VIRCAM's corrective lenses are closer to the focal plane than WFCAM's, reducing relative flexure.
4. The chromatic aberration in VIRCAM is smaller.

Additional effort, such as monitoring trends in VIRCAM-2MASS residuals over many frames and long timescales may be capable of reducing systematics further below the 0.1 arcsec level , but this is outside the scope of the current VIKING plans. Astrometry to 0.1 arcsec systematic error is more than adequate for all our main science aims.

3.3 Photometric calibration

For photometry, the standard instrumental signature removal of Section 3.1 first corrects for dark current, non-linearity, bad pixel masking, flat-field variations etc.

After this, there are in principle three independent routes to photometric calibration:

1. Matching to 2MASS stars, with suitable colour equations.
2. Using the nightly standards (for photometric nights).
3. Global solution using matching of overlapping tiles.

The main photometric calibration will be (a) 2MASS, with method (b) used as a check and (c) applied in the longer term when sufficient overlaps are available.

Each VIRCAM pawprint will contain over 100 useful 2MASS stars (SNR > 10 in 2MASS, and also unsaturated in the VIKING frames), corresponding to e.g. $13 \lesssim J \lesssim 15$.

As an initial one-time procedure, colour equations transforming the 2MASS system to the VIRCAM system will be derived of the form

$$J_t = J_2 + C_J(J_2 - H_2), \qquad H_t = H_2 + C_H(H_2 - K_2), \dots ,$$

where J_2 etc is 2MASS magnitude, J_t is transformed to the VISTA filter system, and the colour terms C_J, C_H, C_K will be derived from fitting to a large number of frames and subsequently held fixed.

For routine reductions, the above colour equations with fixed coefficients give a transformed magnitude J_t, H_t, K_t in the VISTA system for each 2MASS star.

Then, only a single zero-point for each pawprint is needed, e.g.

$$J_{cal} = J_{ins} + ZP_J - e_J(X - 1),$$

where $J_{ins} = -2.5 \log_{10}(ADU/\text{sec})$ is the raw VIRCAM instrumental magnitude, ZP_J is the zeropoint, e_J (normally fixed) is the extinction coefficient, X is airmass and J_{cal} is the calibrated VIRCAM magnitude on a standard system e.g. Vega. Thus, fitting $J_{ins} - e(X - 1)$ vs J_t should give a line of slope 1, intercept ZP_J and small scatter due to 2MASS random errors and colour residuals; both of which average down in the final ZP_J.

This assumes that the 2nd order colour term from 2MASS to VIRCAM magnitude, and the colour-dependence of extinction, are both negligible: these are generally a good approximation in the near-IR where most stars have relatively smooth spectra. Errors in the assumed extinction coefficients cancel to first order since they give an opposite error in ZP (if a per-frame zeropoint is adopted). Also, this method is robust against isolated 2MASS errors, since any single VISTA tile overlaps with a large number of distinct 2MASS stripes.

If a night is photometric, the instrument response is stable and the extinction term is correct, then all frames in the ith passband should give the same value of ZP_i . Analysing trends with time or airmass can reveal non-photometric nights, long-term drift in throughput or gain (e.g. dust accumulation on the optics or IRACE gain drift) or errors in the assumed extinction coefficient.

For Z and Y bands, the situation is slightly more complex since there are no direct 2MASS measurements. (SDSS z-band data is available for the NGP, and z-band data will exist soon in the SGP from the Australian Skymapper project). As a first pass, we intend to use the well-defined stellar locus to bootstrap from 2MASS J,$\mathrm{K_s}$. In a 2-colour diagram such as Y-J, J-$\mathrm{K_s}$, normal stars form a tight sequence with a spread below 0.1 mag rms (given low extinction), so the peak of the stellar locus can be located to better than 0.02 mag. (The locus is not a straight line, but this does not affect the result as long as it is accounted for). We will firstly select a set of fields observed in photometric conditions, and fit the peak of the stellar locus e.g. $(Y-J)_{loc} = f(J-\mathrm{K_s})$. Then, for Z,Y tiles observed in non-photometric conditions, the zero-point can be bootstrapped from the J,$\mathrm{K_s}$ data; e.g. given J,$\mathrm{K_s}$ measurements zeropointed on 2MASS as above, the predicted Y-magnitude (if the star were on the locus) is $Y_{loc} = J_{obs} + f(J-\mathrm{K_s})$. The Y zeropoint can be set to give zero *mean* deviation of all stars in the frame from the locus, thus the intrinsic scatter around the locus averages down over the number of stars. Once our J,$\mathrm{K_s}$ frames are zeropointed from 2MASS, our numerous fainter stars can be used in fitting the Y locus hence random errors from the locus width should decrease to ~ 0.02 mag or less.

The regular $\sim$ 2-hourly touchstone-field observations will be used to derive nightly zero-points (on photometric nights) as an additional check of this procedure.

As a final step in photometric calibration, an **"illumination correction"** will be applied to correct for the fact that a standard flat-fielding procedure does not lead to precise photometry, due to two effects: firstly distortion leads to small variation in pixel areas on the sky ($\approx 3\%$ peak-to-valley), and secondly stray-light forms an *additive* and roughly axisymmetric background offset (see e.g. Ref. [07]). The illumination correction will be a position-dependent offset calculated from either stacked residuals vs 2MASS, and/or mesostepped frames across the touchstone fields.

Based on WFCAM experience, we hope to achieve a photometric zero-point accuracy of ≈ 0.02 mag rms at J,H,$\mathrm{K_s}$ and ≈ 0.03 mag rms at Z,Y.

4 Data reduction process

We will use the VISTA Data Flow System (VDFS, Ref. [03]) for all aspects of data processing and archiving, except for the QC0 + QC1 stages which take place at Paranal (using modules supplied by VDFS). The Cambridge Astronomical Survey Unit (CASU) at Cambridge will be responsible for pipeline processing and first-level calibration, and the Wide Field Astronomy Unit (WFAU) at Edinburgh will be responsible for the Science Archive with interface and query facilities , and creation of higher-level products e.g. bandmerged catalogues, list-driven photometry etc.

For a more detailed description of this system see `http://www.ast.cam.ac.uk/vdfs/` (CASU), and `http://www.roe.ac.uk/~nch/wfcam/` (WFAU) .

Both the VDFS pipeline and archive facilities have been designed specifically for VISTA, and have already been scientifically verified by processing wide-field near-IR imaging from UKIRT's WFCAM imager, at routine rates up to 250GB/night. Versions of the pipeline have also been used to process ESO ISAAC data, and data from a wide range of optical CCD mosaic cameras.

We note here that VIKING, while clearly a large and ambitious survey, is relatively undemanding in VDFS resources compared to many other planned Public Surveys. Since VIKING is near the median of the planned VISTA Public Surveys in terms of depth and area, it does not stress the overall dataflow system in either extreme: it has lower data volumes and less time-dependency than the shallower surveys, while it has less stringent demands for removal of low-level systematic effects compared to the narrower deep surveys e.g. VIDEO and UltraVISTA.

4.1 Pipeline processing

Data will be transported from Paranal to Garching by ESO's standard procedures (currently hard disk) and next to CASU also by hard disk. These 'moving' hard disks are copied to the local disk towers at Garching and Cambridge, then continue on a closed loop back to Garching and Paranal. Checksums are checked, `fitsverify` is run, and failures trigger a semi-automated recovery process. Raw data is also copied onto LTO tape for local backup.

The pipeline (hosted at CASU) will perform all standard processing steps for instrumental signature removal, including

- Dark subtraction, using library dark frames.
- Linearity correction using a linearity calibration table, derived from a series of dome flats of varying exposure time at constant illumination level.
- Bad-pixel masking, using analysis of the dark and linearity frames as above to define a bad-pixel mask (either hot or defective pixels).
- Flat-fielding using twilight sky flats ;
- "De-curtaining". This is removal of a row-to-row bias level variation, which is time-dependent hence not removable with dark frames , (see Ref. [02] for a more detailed explanation).
- Cross-talk correction (if required ; up to now, no significant cross-talk has been observed in VIRCAM).
- Assessing and dealing with image persistence from preceding frames (if necessary and if practical).
- Sky background removal using a stack of images with object rejection and robust averaging to define a local background frame.
- Co-addition of jittered frames for a single pawprint, using sky background removal and cosmic-ray rejection algorithms.
- Bad-pixel handling, propagation of error arrays and effective exposure times by use of confidence maps associated with processed images.
- Generation of single-passband image catalogues.
- Merging of individual stacked pawprints to provide a single tile per passband resampled onto a common linear tangent-plane coordinate system.
- Calibration to a standard photometric system (see Section 3.3).
- Astrometric calibration to ICRS system based on 2MASS with appropriate World Coordinate System information recorded in the FITS headers.
- Processing history, including pipeline software versioning and calibration files, are recorded in the FITS headers.

Library frames are crucial and are built with a significant amount of human checking. Basic quality control (e.g. rejection of frames with serious cloud, tracking/guiding failures, EMC interference, Moon problems if any, other hardware problems) will also be carried out at this stage. This uses automated procedures to derive a set of standardised quality control parameters, then compare these with typical values and generate warnings if they lie outside an approved range (see Section 6 for more details of this process).

These processed and stacked pawprints, with instrumental signatures removed and calibration information, confidence maps and history in the FITS headers, will then be delivered to the VISTA Science Archive at WFAU, Edinburgh.

4.2 Science Archive

The VISTA Science Archive (VSA) at WFAU is an extension of the WFCAM Science Archive, which is currently operational and serving the full set of UKIDSS surveys to the ESO community.

The VSA will ingest the outputs of pipeline processing into a database, and will then curate these to produce enhanced database-driven products.

The functions carried out by VSA will include:

- Individual passband frame association.
- Source association to provide multi-colour source lists including the 2 epochs of J-band.
- List-driven photometry ; this is a key objective for VIKING, since it will be extremely valuable to derive list-driven photometry on the VIKING images for the list of objects selected in KIDS *i*-band. A catalogue with list-driven photometry based on J-band selection will also be provided.
- Quality control features, as defined and led by the VIKING team, supported by the VSA team.
- Generation of new image products, e.g. coadds of our two distinct J epochs.

The VSA will have a user-friendly interface based on SQL queries; both simple and advanced interfaces are available, with the simple interface for ease of use while the advanced interface exposes the full relational database structure to the user enabling more complex queries and manipulation.

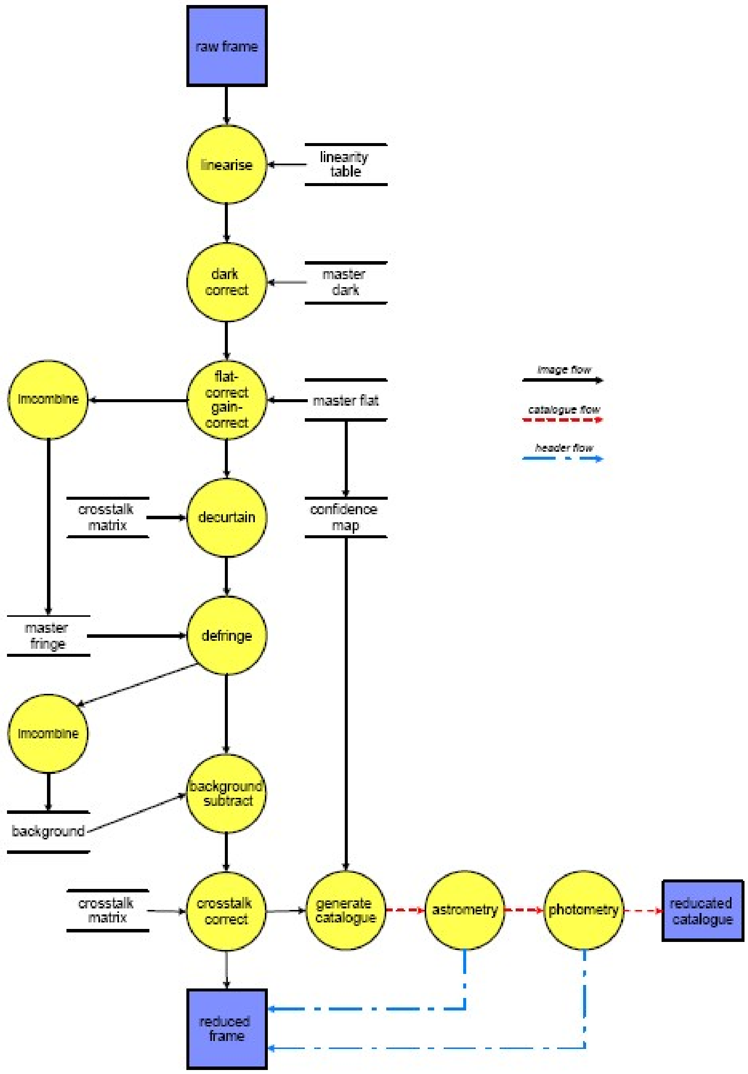

Figure 2: A block diagram outlining the major steps from raw data to the calibrated data products.

5 Manpower and hardware capabilities devoted to data reduction and quality assessment

5.1 Team members:

The main VIKING and VDFS team members with specific responsibility related to data reduction and quality assessment are listed in the following table:

Name	Function	Affiliation	Country	FTE (annual)
W. Sutherland	VIKING PI	QMUL	UK	0.7
J. Emerson	VDFS Manager	QMUL	UK	0.2
M. Irwin	CASU manager	IoA, Cambridge	UK	0.7*
J. Lewis	Pipeline development	IoA, Cambridge	UK	0.8*
M. Riello	Pipeline operations	IoA, Cambridge	UK	0.8*
S. Hodgkin	Photometric calibration	IoA, Cambridge	UK	0.7*
N. Hambly	WFAU manager	IfA, Edinburgh	UK	0.7*
M. Read	Archive operations	IfA, Edinburgh	UK	0.8*
E. Sutorius	Archive operations	IfA, Edinburgh	UK	0.8*
New postdoc	Post-pipeline science QC	QMUL	UK	0.7
J. Findlay	Post-pipeline science QC	QMUL	UK	0.5
K. Kuijken	KIDS PI	Leiden	NL	0.2
E. Valentijn	ASTROWISE coordination	Groningen	NL	0.2
A. Edge	UKIDSS cross-comparison	Durham	UK	0.2
S. Driver	PSF modelling, sky-subtraction QC	St. Andrews	UK	0.2

Note 1: * denotes FTEs shared between VISTA surveys; a substantial fraction of this involves tasks common to many surveys, while a smaller fraction is VIKING-specific.

Note 2: The full set of VIKING co-Is includes many scientists (31 co-Is total) responsible for varied science analyses of the survey data, in addition to those above. The list above is limited to responsibilities directly relating to the OB preparation, data processing, science quality control and delivery of resulting data to ESO.

5.2 VDFS facilities

Based on experience running the WFCAM pipeline, the CASU pipeline will have 3 FTEs available for steady-state operations on VISTA; this includes normal processing, reprocessing after major bug fixes or enhancements, system maintenance and upgrades, and liaison with users. The VISTA pipeline will be a key priority of CASU over this period; the unit is funded via 5-year rolling grants with renewal every few years (see below). The unit will contain a cluster of 16 processors (2.4 GHz Opteron) dedicated to processing, plus additional CPUs for redundancy and data-handling. The hardware requirements are specified to have an overcapacity factor of at least 3 , to allow for inevitable fluctuations in the data rate and reprocessing requirements. Hard disk storage is initially 3×10 TB RAID6 arrays, expandable as required. All raw and processed files will be stored using lossless Rice tile compression to save a factor of 4 in disk storage requirements.

The prime responsibility of WFAU in the next few years will be maintaining the VISTA and WFCAM Science Archives, plus existing archives. Manpower provision at the WFAU archive comprises 2 FTE of dedicated operations staff and 1 FTE of astronomer-scientist management, oversight and system support. Manpower and hardware provision for storage of pipeline-processed science product files, database server catalogue storage, web servers and other infrastructure, is funded via the UK Science and Technology Facilities Council (STFC, formerly PPARC).

Both the VDFS components at CASU and WFAU are funded from STFC rolling grants. The next periodic renewal has a submission date of 5 June 2007, and will provide funding for the period April 2008 – March 2013. The request being made will take account of the needs of VISTA Public Surveys including VIKING, and the above lifetime of the renewed grant covers the anticipated 5-year operating period of the VIKING survey.

We are confident that the hardware will be adequate for VIKING since the CASU+WFAU system is already proven to handle the data rate from UKIDSS. Its processing and storage capacity is being upgraded by $\sim 3\times$ for VISTA, driven by the needs of wider surveys such as VISTA Hemisphere Survey and VISTA Variables in the Via Lactea (VVV), which generate the highest raw data flow rate. In fact, for VIKING the average raw data rate is only slightly higher than that for the UKIDSS Large Area Survey. Compared to the latter, VIKING has $\sim 1.5\times$ more hours/year, and $4\times$ more pixels, but VIKING stores fewer frames per unit time, typically around

1/4 of the frames per unit time due to coadditions within IRACE. Therefore, given the substantial expansion of hardware already planned, CPU and disk resources should not normally be a significant bottleneck for VIKING.

5.3 VIKING team

The VIKING team will be responsible for overall testing that the processed data (after passing Paranal QC1 and the standard set of CASU pipeline tests) meets the final science requirements as far as reasonably possible.

The PI will take overall responsibility for data quality; as VISTA Project Scientist he has a detailed knowledge of the telescope, instrument and the VDFS system, and also has extensive experience with large extragalactic imaging and spectroscopic surveys including the APM Galaxy Survey, the MACHO project and 2dFGRS. A new PDRA and PhD student under the supervision of the PI will perform the regular checking beyond the CASU automated tests, e.g. inspection of colour magnitude diagrams and visual checking of potential anomalies. Emerson (VDFS PI) will be responsible for overall coordination between the VIKING project and the VDFS. Kuijken (KIDS PI) will have responsibility for the KIDS coordination and prioritisation of the relevant sky areas, while Valentijn will have responsibility for the interface between Astro-WISE and the VISTA Science Archive. Edge is a core member of the UKIDSS science team and will lead the VIKING-UKIDSS cross comparison in the NGP stripe. Driver is expert in galaxy populations, and will perform detailed modelling of the PSFs, background subtraction and galaxy detection completeness as part of these studies.

The tasks planned are listed in more detail in Section 6.3.

6 Data quality assessment process

The assessment of data quality essentially is a 3-stage process: a quick assessment (QC0, QC1) is performed at Paranal, then a second stage is performed at VDFS, and the final stage before general data release is the responsibility of the VIKING team. These stages are outlined in more detail below.

6.1 Paranal quality control

An initial QC1 stage is carried out by modules supplied by VDFS and running on a dedicated multi-processor workstation at Paranal, for near real-time feedback to the telescope operator. The Paranal pipeline is required to be "causal" (i.e. unlike the full VDFS pipeline, any frame must be able to be reduced immediately without dependence on future frames later in its OB or night), so this is only used for health-checks, providing daily feedback to the Paranal staff. Therefore, this is separate from the main data reduction pipeline, though the same QC parameters (and additional ones) will be re-generated as part of the UK reduction process.

6.2 VDFS quality control

In the VDFS pipeline, considerable effort has gone into the design of automated QC parameters which are generated automatically during the reductions and compared with typical ranges. All of the CASU pipeline QC information will be made available to the VIKING team via a web-based QC interface (for an example, see Ref. [05]) and these QC parameters are also recorded in the FITS headers.

A complete list of QC parameters are available in Ref. [02] Appendix A , while some examples include:

- Pointing differences between blind telescope pointing and final calibrated WCS.
- Mean sky level and rms noise.
- Number of objects classified as "noise".

- Saturation level (from bright stars)
- Mean FWHM and ellipticity for objects classified as "stellar".
- Aperture correction (fraction of stellar flux outside 2 arcsec).
- Photometric zeropoint derived from 2MASS matching.
- Photometric zeropoint derived from nightly touchstone fields.
- Stellar magnitude limit (computed from the above).

6.3 VIKING team quality control

Additional tests will be done at the post-archive stage by the VIKING QC team, including the following:

- Additional checks of QC parameters generated by the pipeline, to look for low-level effects or trends which may not trigger the automated warnings.
- Inspection of colour-colour and colour-magnitude diagrams. In general stars and galaxies lie on distinct sequences in near-IR colour-colour space, with most stars forming a relatively tight locus, so inspection of these plots plus morphological classification forms a powerful check for many systematic errors (see Ref [06] for examples).
- Inspection of object plots: Localised image problems are usually apparent by inspection of object dot-plots, especially for objects of unusual colours. Problems creating significant numbers of spurious images (such as diffraction spikes, aircraft or satellite trails, bright-star ghosts, etc) are readily apparent. Also, since stars at high latitudes are very weakly clustered, comparison of stellar vs galaxy dot-plots forms a good test for spurious images or breakdown in star-galaxy classification.

  Based on early experience, we will create an automated masking around bright stars, i.e. drill a magnitude-dependent radius around each bright star. Aircraft or satellite tracks are clearly unpredictable, but show up as a linear track of single-band detections hence are readily detectable by a visual inspection.
- Overlap matching. The VIKING field pattern will provide significant edge overlaps ~ 2 arcmin wide at the North and South edges of each tile, containing > 500 objects per overlap. Thus, comparison of image parameters for objects duplicated in the overlap regions will provide a strong check of most systematic errors (with the exception of centre-to-edge field-dependent systematics, but the latter can be tested by stacking residuals vs 2MASS).

  Larger samples of overlap objects are available from the 'wings' 5.5 arcmin wide at the North/South overlaps, which receive full exposure in one tile and half the normal exposure in the other.
- KIDS matching. Our survey is intended to be a 9-band optical-NIR survey, thus matching with KIDS is a very important element. This will occur as part of the WFAU archive ingestion, and this will clearly provide an additional quality-control step based on comparisons of the independent astrometry from KIDS + VIKING, and comparing stellar loci in multicolour space.

  As noted above, KIDS and VIKING tile centres will match in the Dec direction but have quasi-random offsets in the RA direction, so each VIKING tile will have a large overlap with either 2 or 3 distinct KIDS tiles, plus additional smaller overlaps at the N+S edges, providing good interlocking to localise potential problems.

7 Data product and VO compliance:

Our fundamental data products comprise the deliverables in the VISTA Science Archive, and will include :

- Image data stored as full tiles, separately for each passband, projected onto a simple projection e.g. Tan projection, and astrometrically calibrated to ICRS with systematic errors below 0.1 arcsec rms. These will be on a standard zeropoint, typically Vega scale, in the VISTA-VIRCAM natural instrumental system.

  Images will be multiextension FITS files with extensive metadata propagated via the image headers, including processing history and software versions etc.

- Confidence maps used and generated during the production of the above images, including automated and/or manual masking of defective sub-regions.
- Single-band catalogues separately for each VIKING passband, including discrete and continuous star-galaxy classification parameters, aperture magnitudes in log-spaced apertures of $\sqrt{2}$ ratio, etc.
- Merged multi-band catalogues created from the full set of single-band catalogues in each tile.
- Catalogue containing list-driven photometry in matched apertures in Z,Y,J,H,K_s for all objects selected and defined in J-band.

Additional products, notably list-driven photometry with multicolour near-IR fluxes and upper limits for a complete list of objects to the KIDS 5σ i-band limit, will normally be made available via the Virtual Observatory-linked VISTA Science Archive at WFAU: however as these products are reliant on availability of non-VIKING data, these may not necessarily follow the standard data delivery schedule. As KIDS or other suitable visible catalogues become available, we anticipate production of matched-aperture catalogues from VIKING data with delivery to ESO in suitable phases to be agreed.

The VSA is being designed for full compliance with the emerging Virtual Observatory standards, including use of self-documenting FITS files, source catalogues as binary FITS tables with full information in header fields, tables in XML format, etc. This will be progressed further as VO standards develop.

8 Timeline delivery of data products to the ESO archive:

Raw VISTA data is normally expected to arrive in the UK roughly 17 days after the observations are taken, in steady state. The turnaround time for extracting and verifying data is expected to be another week assuming no significant problems.

In the steady state operations, CASU will normally complete pipeline processing within approximately two months after the data are delivered; we aim to perform the VIKING QC within another two months, allowing margin to either mask out or re-reduce isolated problem areas before ESO delivery.

Therefore, during the steady state, VDFS will normally be able to deliver the relevant survey data products to the ESO Science Archive before the end of the **semester following** the one in which the raw data were delivered to the UK .

We anticipate that a longer delivery period will be required for the first one or two semesters' worth of data, in order to make provision for a more extended quality control and analysis period by the VIKING team, and if necessary a reprocessing phase with improved software parameters to correct problems discovered in the first-look analyses. In order to allow for this, in practice the reprocessing is likely to start during the third semester. We estimate that the data products from the **first two** semesters of VIKING data should be delivered to ESO not more than one year later than the end of the second semester, therefore at the **end of the 4th semester**, counting the 1st as the first semester in which a significant quantity ($>$ 50 hours) of VIKING data is taken.

Thus for example, assuming VIKING begins in period 80, the delivery of periods 80+81 data would occur at the end of period 83. This will clearly be somewhat in advance of a presumed 2-year review by the Public Surveys Panel.

Data from the 3rd and 4th semesters will be delivered to ESO at the end of the 5th semester, thus bringing the 4th semester into line with the steady state, and thereafter we will continue with the normal one-semester delivery timescale.

In order to deliver some data earlier than the above 4th semester, we suggest that some early Science Verification data be taken early; this could form the basis for an Early Data Release at the end of the 2nd semester, with suitable caveats e.g. to lower quality standards than the regular deliveries above.

It is currently planned that data will be delivered from VSA to the ESO Science Archive via a high-speed Internet connection.

9 References

[01] P. Bunclark and S. Hodgkin. "VISTA Infrared Camera Calibration Plan", VIS-SPE-IOA-20000-0002, v1.4, 2006/11/13.

[02] J. Lewis, P. Bunclark, S. Hodgkin., "VISTA Data Reduction Library Design", VIS-SPE-IOA-20000-0010, v1.6, 2006/12/20.

[03] J. Emerson, M. Irwin, N. Hambly, 2006, "VISTA Data Flow System: Status", Proc. SPIE, 6270, 30.

[04] http://www.vista.ac.uk/vdfs/esoqc1/

[05] http://casu.ast.cam.ac.uk/surveys-projects/wfcam/data-processing/

[06] Dye, S. et al, 2006, MNRAS, 372, 1227.

[07] Manfroid, J., Selman, F., Jones, H., 2001. ESO Messenger, 104, 16.