2MASS PRODUCTION PROCESSING SYSTEM (2MAPPS)

FUNCTIONAL REQUIREMENTS (VERSION 1.2)

TABLE OF CONTENTS

LIST OF TABLES

I. INTRODUCTION AND SUMMARY

The Two Micron All-Sky Survey (2MASS) will employ two ground-based telescopes to scan the northern and southern skies in three near infrared wavelength bands. The images produced by these scans will be used to conduct a census of point sources and galaxies that is nearly 80,000 times more sensitive than that produced by the only previous large-area near infrared survey. A brief overview of the 2MASS project and data products are provided in Tables I and II.

The processing of the raw survey image data from 2MASS will be conducted at the Infrared Processing and Analysis Center (IPAC). This document describes the Scientific Functional Requirements on the Two Micron All-Sky Survey Production Processing System (2MAPPS). These requirements are derived from the 2MASS Level 1 Science Objectives and Specifications, summarized in Table III, and the October 1994 2MASS Proposal. In this document, each stage of data processing that is required to satisfy these objectives is defined, from receipt of data at IPAC to the production of the final 2MASS data products.

Table I. 2MASS Overview

Wavelengths	J(1.25 μm), H(1.65 μm), K_s(2.17 μm)
Sky Coverage	$\geq$ 95% for $\lvert b \rvert < 10^o$
	no gaps $\geq$ 200 sq.deg. all sky
Telescopes	two 1.3-m (Mt. Hopkins, CTIO)
Detectors	3 NICMOS III arrays (256x256)
Camera Pixel Size	$2.0''$
Field of View	$8.5' \times 8.5'$
Integration time	7.8 sec

Table II. 2MASS Data Products

Image Atlas	Photometrically calibrated FITS format images with 1.0″pixels in three wavelength bands. J, H and $K_s$images put onto a common coordinate system for easy comparison.
Point Source Catalog	Catalog containing ≥100 million band-merged unresolved sources (Table IV).
Extended Source Catalog	Catalog containing ≥0.5 million band-merged galaxies and small nebula (Table V).
Explanatory Supplement to 2MASS products	
Archive of raw data and tools for their analysis	

Table III. Image Atlas Requirements

Category	Requirement
Pixel Size	$\frac{1}{2}$(raw camera pixel) $\sim 1.0''$
Background Correction	Match medians of frames before coadding frames along a single scan
Photometric Calibration[1]	Adopted from Point Source Catalog
Astrometric Calibration[1]	Adopted from Point Source Catalog
Distortion	≤0.1% edge-to-edge of one frame

[1]For areas that do not include dead pixels.

A. Relationship To Other Documents

As described above, this document defines functional requirements for 2MASS data processing that are derived directly from the Level 1 Science Objectives and Specifications. As yet, a number of parameters describing survey strategy and geometry remain tentative, so when necessary, this document defaults to the survey description provided in the 2MASS Proposal submitted to the National Aeronautics and Space Administration (NASA), dated October 3, 1994. The data received by IPAC from the 2MASS Observatories will have

Table IV. Point Source Catalog Requirements

Category	Requirement
3-Waveband Sky Coverage at $\|b< 10^o$	$\geq$95%
Overall Coverage	no gaps $\geq$200 sq.deg.
Sensitivity (SNR$\geq$10)[1]	$K_s \leq$14.3 mag.
	H$\leq$15.1 mag.[3]
	J$\leq$15.8 mag.
Photometric Uniformity[2]	$\leq$4%
Photometric Precision[2]	$\leq$5%
Brightest Measurable Stars:	
Photometric Bias	$\leq$2% for $K_s \geq$4.0 mag.
Repeatibility	$\leq$5% for $K_s \geq$8.0 mag.
	up to 10% for $K_s \geq$4.0 mag.
Position Reconstruction[4,5]	0.5″
Differential Reliability of Single Band Sources[6]	$\geq$0.9995
Differential Completeness at $\|b\| > 30^o$ [6]	>0.99

[1] Refers only to sources with $|b| > 10^o$. Below this latitude, confusion and interstellar extinction are likely to affect the sensitivity and accuracy of the survey.

[2] 1σ

[3] The feasibility of this requirement is uncertain due to lack of knowledge of the background variations at H.

[4] Refers only to sources detected at SNR$\geq$20.

[5] The absolute accuracy will depend on the reference star catalog; at present, The Guide Star Catalog is the best reference having the required surface density and sky coverage.

[6] Per mag. in the faintest mag. bin. Applies only to flux levels brighter than the confusion limit.

a form and media defined in the 2MASS Observatory/IPAC Tape Interface Specification Document, Rev. G.

Table V. Extended Source Catalog Requirements

Category	Requirement[1]
Sky Coverage at $\lvert b \rvert > 10^o$	$\geq$95%
Overall Coverage	no gaps $\geq$200 sq.deg.
Sensitivity	$K_s \leq 13.5$ mag.
	$H \leq 14.2$ mag.
	$J \leq 15.0$ mag.
Photometric Uniformity[2]	$\leq$10%
Photometric Precision[2]	$\leq$10%
Differential Reliability[3,4]	$\geq$0.80 ($\lvert b \rvert \geq 10^o$)
	$\geq$0.99 ($\lvert b \rvert \geq 20^o$)
Differential Completeness[3,4]	>0.90 ($\lvert b \rvert \geq 30^o$)

[1] Refers only to sources with $\lvert b \rvert > 10^o$. Below this latitude, confusion and interstellar extinction are likely to affect the sensitivity and accuracy of the survey.

[2] 1σ

[3] Galaxies with brightness greater than the above limits, and $\alpha \geq 0.5''$, where the intensity is modeled as $I = I_{central} e^{(-r/\alpha)^{1/\beta}}$, and $\beta \sim 1$.

[4] Refers only to sources detected at peak SNR$\geq$20.

II. INPUT PROCESSING

For each successful night of observing at the observatory, approximately 20-25 GBytes of quality assured data will be delivered to IPAC as a series of 8mm tapes, each holding ~5 Gbytes of data. Data from each wavelength will be stored on separate tapes. The data on each tape will have been compressed at the observatory by approximately a factor of 2 using a lossless compression algorithm (TBD). Each tape will have as its first file, a Tape Directory file listing the complete contents of that tape, as specified in the Observatory/IPAC Tape Interface Specification.

The first tape read at IPAC as part of the Input Processing for each night will be the last tape in each data set. In addition to the Tape Directory file, this tape will contain two Log files that describe the data contained on the other tapes. The first of these, the Observatory Scan Log file, will provide the 2MAPPS system with information that uniquely identifies each scan taken that night, locates that scan's data in the set of tapes, and contains all relevant parameters that describe that scan necessary for processing, including starting equatorial position, starting UT, direction and length. The second file is the Observer's Log file, and will list the operator's quantitative assessment of the weather conditions and any possible anomalies associated with each scan.

IPAC will perform initial integrity checking on each data tape by reading that tape, decompressing the data and logging each scan as soon as the data arrive at IPAC. Each data set read from a tape will be compared with the Tape Directory File and the Observatory Scan Log file to verify that all data for that night is present. Tape problems or missing data will be communicated to the 2MASS project office within 48 hours of receipt at IPAC.

Scans that were taken in poor weather conditions or that have other anomalies (TBD), as reported in the Observer's Log, will be flagged and rejected for processing, although all raw data will be archived (§XIV).

A. Data Types

The data that will be processed by IPAC will be of two types:

1. **Scan Data -** The science data that will be taken with the telescope, chopping secondary and integration time synchronized to nominal survey values. The header will distinguish two types of scans: calibration scans (CAL) and survey (SURVEY) scans. Calibration scans will be at least 2 (TBD) degrees in length targetting standard stars. Survey scans will be no shorter than 6 (TBD) degrees in length.

2. **Dark Data -** Dark images are obtained with the cold shutter closed and the telescope stationary and are used to correct the on-sky Scan Data for the level and pattern of the detector's electronic bias and dark-current. Two hundred (TBD) dark frames will be taken at the nominal survey integration time. Both $\text{Read}_2 - \text{Read}_1$ and Read_1 frames will be recorded.

Issues:

Content and format of Observer's Log file
Array responsivity test observations - CAL data?
Weather rejection criteria

III. ARRAY CALIBRATION

Array calibration processing will remove the signature of electronic bias in each image (dark subtraction), determine the relative responsivity of each pixel in the detector arrays (flat-fielding), and correct for the presence of known bad pixels in each detector array. The residual errors in photometry due to uncertainties in array calibration shall be less than TBD%.

A. Assumptions

It is assumed that the Observatory will provide data of sufficient quality and quantity that dark correction and array responsivity can be accurately determined. It is also assumed that any structured noise due to electrical interference will have an amplitude less than 0.1% of the J sky brightness per pixel (0.5 DN), and less than 0.01% (TBD) of the J sky brightness averaged over N$\sim$ 100 pixels.

B. Bad Pixel Mask

Template images describing the locations of known bad pixels in each survey camera detector will be provided by the project. Known bad pixels are defined as having very poor (<TBD%) or variable (>TBD%) quantum efficiency, as determined from laboratory tests or test sky data. The template image, or mask, will have pixel values of 1 in all good pixels and 0 in all bad pixels.

C. Dark subtraction

For each night's data, the average of all the dark frames will be created for each band for both Read_2–Read_1 and Read_1 data. Suspect pixels will be identified in the average dark images as having large standard deviations. The average dark frame will be logged and compared with previous dark frames in the history log. Significant (TBD) differences in the number of suspect pixels, or deviations in the dark values will be written out as part of the Q/A process, and the 2MASS project office will be notified. The appropriate dark frame will be subtracted from all scan frames.

D. Flat field generation

In principle, the best map of array detector response is made using data with the same illumination level, color and structure as the images to be corrected. Therefore, detector response maps, or "flats", will be computed for each scan by combining frames within the scan itself using a trimmed-averaging algorithm. The number of frames to be used to generate each flat, and thus the number of flats for each scan, will be set by the requirement that the noise introduced by errors in the flat be less than the background photon noise. The number of flats and the image stack size may be varied. A variety of quality checks will be performed to ensure that the flat is not contaminated by residuals from astronomical sources or by temporal variations in the sky/thermal background.

The quality of the flats will be such that the flattening process will introduce no structure in excess of $\sim$ 0.1% of the sky brightness in a flattened frame. It is recognized that this level of performance may be difficult to meet in the J band if the survey exposure time is shorter than necessary to make the scan frames photon-noise limited, and in regions of very high source density. Alternate (TBD) methods of flat fielding may be employed in these cases.

The resultant flats will be logged for future reference. Significant differences in the number of bad pixels, or deviations in the pixel values will be written out as part of the Q/A process. The output of the flat field process will be used to update the bad pixel mask by turning off pixels deviating by more than TBD sigma from the average responsivity. The average value of the flat will be logged as a monitor of array performance and sky conditions.

E. Array Signature Removal

Each frame of scan data will be masked using the bad pixel template image, dark-subtracted and flattened using data obtained as outlined above.

F. Processing of Read_1 Data

Similar calibration steps will be applied to the Read_1 frames.

Issues:

Alternative flat-fielding methods (i.e. twilight sky, all night's data)
Flat-fielding in high source density regions
Flat-fielding if J data not photon noise limited
Flat-fielding Read_1 images

IV. POINT SOURCE PROCESSING

A. Assumptions

1. It is assumed that the telescope scanning motion will not be stable enough to allow use of the *a priori* frame-to-frame offsets for image alignment and position reconstruction. Therefore, offsets will need to be determined from the images themselves in an initial processing step. To meet the Level 1 positional requirements for all three bands, it will be necessary to determine the frame-to-frame offset using combined J, H, and K_s images.

2. The number of frames per celestial position, NFRAME, is assumed to be 6 to ensure adequate sampling and bad pixel rejection. The minimum integration time will be 1 second to avoid freezing the speckle pattern at K_s. In what follows we assume an integration time of 1.3 seconds per frame.

3. The performance of the optical system has a strong effect on the complexity of the software design and on the eventual number of operations that must be carried out per pixel. The stringent design goals given below (Table VI) will minimize software development and operations costs. A looser requirement for optical performance will have moderate impact on software development and operations costs:

- Improving the encircled energy performance of the system will increase point source sensitivity and make it easier to distinguish galaxies from stars.
- The plate scales should be the same at J, H and K_s within 2% since the smallest field of view will determine the survey coverage.
- Negligible optical distortion will simplify the generation of images (§XI), the position reconstruction of point sources, and the process of making a PSF representative of a particular point source seen on NFRAME different parts of the array.
- Each array will be rotated from the scan direction by an angle of $\tan^{-1}(1/256) = 0.22°$ so that each sighting of a source falls onto a different part of a pixel as its image progresses along the array. Smaller rotation angles would exacerbate problems with bad rows and introduce serious completeness issues.

Table VI. Optical Requirements

Parameter	Requirement	Comment
Encircled energy	90%	within 1 pixel
PSF variation over field	$0.2''$	FWHM over full field
Band-to-band image scale	0.5%	
Band-to-band optical axis	5	pixels
Distortion over field	$0.2''$	
Rotation Angle	0.2 ± 0.02	deg, all three bands
Focus Stability	20%/hour	width of 90% Enc. Energy

- More stringent rotation and plate scale goals ensure uniform scanning of the stellar image across the pixel response function and would simplify band-to-band registration.
- The focus should be stable at least on the time between calibration observations (0.5-1.0 hr) and preferably over the entire night. TBD: A focusing sequence should be built in to every calibration observation.

B. Point Spread Function Determination

A combination of optical analysis using the 2MASS optical prescription and data obtained under good seeing conditions during the initial checkout period will be used to determine the variation of the Point Spread Function across each array. These "perfect" PSFs will be used in conjunction with observations of bright sources on each scan to determine the PSF appropriate for use on that scan by the point source photometry processor.

C. Point Source Detection Requirements

1. Point sources will be detected using an optimized algorithm that identifies all excursions more than 4σ (TBD) above the local noise in the average of the NFRAME frames.

2. A parameter estimation algorithm will generate source characteristics for each detection, including centroid location and uncertainties, intensity (integrated/peak), goodness of fit to PSF profile, and relative photometric uncertainty. The uncertainty should reflect the effect of neighboring sources, or sources in the region selected for background estimation. The estimation algorithm will take into account frames with bad pixels within TBD of the computed position of the source on that frame. The processor will use the locations of the sightings of a source across the focal plane to determine the appropriate PSF.

3. The estimation algorithm will have special logic for estimating the parameters of multiple stars and should be able to characterize accurately all binary stars separated by

more than two camera pixels, and to identify as potentially confused all sources of roughly equal magnitude separated by one camera pixel.

4. A post processor will remove from the source lists all objects due to persistence in the NICMOS array and ghost images from bright sources. The brightness of the initial source will determine the distance and brightness criteria to be used in rejecting objects.

5. There is a goal to improve the photometric accuracy on bright sources (<2% intrinsic uncertainty due to photon statistics). Any change in estimation procedure between faint and bright sources should be carried out as a smooth function of source brightness, and should introduce no systematic differences larger than 1

D. Bright Source Characterization

The same processing steps will be applied to the Read_1 data to estimate the parameters of all sources brighter than TBD at J, H, and K_s. This method should give a ~3.5 magnitude increase in dynamic range relative to Read_2–Read_1 data. The consistency of the brightness measurements between Read_1 and Read_2–Read_1 data will be monitored to ensure photometric continuity over the entire range of magnitudes. Finally, a curve of growth analysis using the Read_1 data may represent a method for measuring extremely bright sources. This technique will be examined during the development phase and incorporated if it proves useful.

A post-processor will merge the source lists from the bright and faint source processors.

E. Performance in Confused Areas

The performance of the point source extractor in confused regions with source densities $\leq$ 100 beams/source shall not be a performance driver on the design of the point source processor. The performance requirements of Table III shall be applicable at a noise floor of K_s=10 (TBD), instead of K_s=14.3, in regions with less than 100 beams/source.

Issues:

Confusion Flags

Implications of band-to-band optical axis offsets

V. EXTENDED SOURCE PROCESSING

A. Assumptions

1. The optical requirements of §IV and Table VI are also critical to meeting the goals of extracting extended sources from the data.

2. To ensure uniform detection limits across the sky, the survey data will be marked as unacceptable for extended source extractions if the sky brightness exceeds TBD limits in J, H or K_s, or if the seeing varies by TBD limits within a scan.

3. The scan-to-scan overlap will be approximately $0.8'$. The galaxy catalog will be complete only for galaxies smaller than this size. Galaxies larger than this size, but smaller than $8'$, will be recovered as long as they do not touch the boundaries of the scan. In

addition, known galaxies with optical radii larger than $1'$ will be flagged in the image data, and the image data will be saved in an extended source image database.

Further requirements for galaxies larger than $0.8'$ would lead to significant complications in the pipeline processing because of the need to combine data from different scans. Larger galaxies can be recovered later by a post-processor for specific galaxies of interest.

B. Basic Processing

The extended source processing shall be optimized to extract information on relatively compact ($< 0.8'$) galaxies in relatively unconfused ($< 10^4$ source/sq. deg.) sky. The algorithm development will not be driven by requirements to find diffuse galactic nebulosity or to operate in highly confused regions.

There are three key aspects of extended source (galaxy) processing: detection, star-galaxy discrimination, and parameter estimation.

1. Critical to the successful detection of low surface brightness emission is the quality of the image data. The limiting detectable surface brightness over a $0.8'$ diameter region should be set by the sky brightness, leading to the strict requirements on the quality of the flattening described in §III.D. Additionally, low surface brightness objects spread out over many pixels must be detected even if no individual pixel lies above a TBD threshold.

2. Galaxies need to be distinguished from stars over as much of the sky as possible. As always, there will be a tradeoff between completeness and reliability in the identification of galaxies.

The processor will use parameters such as central surface brightness and angular size to distinguish galaxies from single and multiple stars with the completeness and reliability requirements given in Table V. The processor will remove artifacts due to bright stars (NICMOS persistence and geometrically well defined ghosts).

3. The fundamental parameters reported for each galaxy will be its position, integrated magnitude or upper limit in each band, and size and shape information. This is all the fundamental information available for the vast majority of galaxies, which are only slightly larger than point sources. Further information for the small number of larger galaxies, such as isophotal magnitudes at various levels, can be computed from the extended source image database.

The extended source processor will report the probability that the detected object is extended, as well as whether the object was also detected by the point source processor.

Other important considerations in the development of the galaxy processor include: guaranteeing that the absolute photometric accuracy for galaxies, averaged over many galaxies in one region of sky, is better than 5%, when compared to another group of galaxies measured anywhere else in the sky; and identifying star-galaxy asterisms and correcting galaxy magnitudes appropriately.

Issues:

Effect of seeing variations on completeness and reliability
Completeness and Reliability as a function of source density
How best to detect low surface brightness galaxies

How to handle stars projected against galaxies
How to handle "merged" galaxies with several nuclei or "touching" or "overlapping" galaxies.
Algorithms to reject triple stellar systems

VI. BAND MERGING

Data from the three bands within a scan (including results of bright source processing) will be used to generate a list of band-merged point and extended sources. Band-merging will be based solely on spatial coincidence, and the spatial windows will be selected to maximize valid band-merges while minimizing artificial merges. Position refinement will use a noise weighted average of the positions in the various bands. Noise fills for bands without detections will come from a noise estimate determined from examination of the data in the coadded image and from the actual measurements of flux in the coadd at the position of the source. The existence of multiple candidates will trigger confusion processing that will try to establish the optimum match based on source brightness and position differences. Appropriate confusion flags will be set.

Issues:

Confusion Flags
Band merging confused sources

VII. POSITION RECONSTRUCTION

A. Assumptions

1. The starting equatorial position of the telescope for each scan will be included in the raw image headers, and will be accurate to within $1'$ ($2\ \sigma$) after correction for any repeatable start-up transients.

B. Processing

The position reconstruction must meet the requirement of $0.5''$ ($1\ \sigma$). If the frame-to-frame offsets can be computed accurately enough to meet this requirement, then a high precision astrometric catalog such as the PPM shall be used to establish the absolute reference system. If necessary, however, position reconstruction processing will match 2MASS point and source extractions to optical reference stars taken from the Space Telescope Guide Star Catalog, Version 1.2, or a higher precision catalog when it becomes available. The spatial fitting procedure will take into account all relevant telescope and camera optical distortions to the degree necessary to meet the position requirements listed in Table II. Information from all three bands will be used in deriving the final pointing solution.

Position reconstruction will provide all the information needed to evaluate positions of point sources, extended objects and all 2MASS images. Uncertainties will be quoted for each source that take into account detection uncertainties, reconstruction uncertainties, and absolute uncertainties in the reference frame. Trend analysis will be used to monitor long term drifts in telescope performance.

Issues: Availability of Hipparcos/Tycho stars for QA

VIII. CATALOG ASSOCIATION

2MASS sources will be compared with a small set (TBD) of stellar and extra-galactic catalogs as part of the main processing pipeline. The association with a broader set of catalogs will be carried out during the final product generation phase and is not considered further here. The purpose of these associations is to identify sources with extreme properties (extremely faint optical counterparts, etc) that might be either processing artifacts or interesting astronomical objects. The association task will examine the catalogs and retain the information described in Table VII.

The associations processor will output information based on the stellar matches to monitor pointing reconstruction.

Table VII. Association Information

Catalog	Information
GSC/Tycho	magnitude, type
APM/Cosmos	magnitude, type
IRAS	pointer to FSC, PSC
VLA	flux density
NED	pointer to NED

Issues:

Do all catalog association as part of production?
Other Catalogs for Association

IX. PHOTOMETRIC CALIBRATION

A. Assumptions

1. The project will provide a list of standard calibration stars spread uniformly across the sky with precisely known positions and magnitudes. The standard stars should fall within the brightness range given in Table VIII to ensure a signal to noise of at least 100 on a scan, but to prevent any non-linearities from affecting the data. A subset of the standards should be at declinations such that they approach airmasses close to 1.0 in either the Northern or Southern hemisphere. This is necessary for determination and refinement of extinction coefficients. It is also very important to have a significant fraction

of the standards near the equator, so that they may be observed from both observatories at low airmass and thus provide the critical cross-calibration of the two halves of the survey.

2. All standards will be observed in TBD-degree scans. In addition to providing more information for tracking flats and PSF variations, a series of secondary standards can be set up and used for photometric verification of the primary standards and the nightly calibration.

3. The survey strategy will ensure that one or two calibrators are observed each TBD hours. The calibration strategy should attempt to make repeated observations of the same calibration fields through the night to build up information on the airmass corrections.

4. The project will provide a list of fields containing stars with a broad range of colors to monitor color terms in the calibration (J-H) from 0 to 2 mag, H-K_s from 0 to 4 mag. These fields will be observed once per night (TBD).

5. The temperature of the arrays will be held stable enough that the variation of responsivity is less than 0.5% over the period between calibration scans.

6. The observatory will provide data of sufficient quality and quantity for each night of observing that allows photometric calibration of the data for that night to be made that satisfies the requirements in Tables III and IV.

Table VIII. Calibration Star Requirements

Band	Magnitude Range (TBD)
J	10-13.5
H	9-12.5
K_s	8-11.5

B. Processing

Photometric calibration will be carried out on a nightly basis and represents the only part of the pipeline that puts a human into the processing loop.

The CAL scans will be processed as part of the regular 2MASS pipeline. The standard stars will be extracted and placed in a file for analysis to determine the nightly zero-point, extinction correction and color terms for each band. Other stars in the calibration fields may be used as secondary standards to monitor extinction variations.

The overall photometric quality of each night will be gauged by the dispersion in the solution to the photometric transformations for each night. Transformations with $\geq$TBD% residuals will be grounds for rejecting data for that night.

Since atmospheric transparency can vary on short time scales, the scan-to-scan overlap sources will be monitored to diagnose short-term variations in transparency and thus identify photometric problems. The processor will use the short-term variations measured in the overlap regions to verify that the block of data between calibration scans was photometrically stable. Scan-to-scan deviations greater than 0.01 mag (TBD) in any band will be grounds for rejecting a block of data.

As much as possible, verification of the calibration of the night's data will be automated. The overall calibration will be certified by the IPAC cognizant scientist using criteria developed in conjunction with the 2MASS science team. Once the calibration is determined, a post-processor will apply the relevant factors to the point sources and to the image headers. The photometric uncertainties will be updated to reflect the absolute calibration uncertainties.

The processor will provide long term diagnostics on weekly and longer time scales for extinction, color terms, and trends. During the protocamera observations and the initial checkout period a large fraction of time should be spent on calibration objects with large color ranges and air masses beyond what will be used in the survey, so we can determine the true requirements for calibration observations.

C. Calibration Working Survey Data Base

Observations of stars in the calibration fields offer a variety of short and long term diagnostics that can be used to verify positional and photometric accuracy, as well as completeness and reliability. These extractions will be put into a Calibration Source database for various quality assurance purposes.

X. WORKING SURVEY DATA BASE

The Working Survey Data Base (WSDB) represents the final repository for all 2MASS sources and must be readily accessible for technical and scientific analysis. Possible fields of a WSDB record are given in Table IX.

Sources will be merged into the WSDB after the positional and photometric calibration data have been verified and applied. The WSDB will be updated and reindexed regularly TBD.

The scan-to-scan overlap will result in duplicate sightings of the same sources. A confirmation processor will characterize the overlap sources in terms of position, photometric stability, completeness and reliability. Duplicate sources will be flagged. The question of what to do with duplicate sources in the overlap regions (merge, maintain separate entries...) is TBD.

Requirements on access to data into the WSDB include: 1) retrieval of 1000 sources based on queries of indexed fields within 1 (TBD) seconds; 2) search of entire WSDB based on a query involving non-indexed fields within 30 minutes (TBD).

XI. SURVEY IMAGES

A. Assumptions

The optical performance described in §IV and Table V is critical to the generation of high quality images.

B. Image Specification

1. An image product will be produced for each band of each declination scan.

Table IX. Illustrative WSDB Record

Parameter	Comment	Parameter	Comment
α^1	band average	δ^1	band average
$\sigma(\alpha)$	uncertainty	$\sigma(\delta)$	uncertainty
Gal Lat^1		Gal Long^1	
J^1	mag/upper limit	σ_J	uncertainty/flag
H^1	mag/upper limit	σ_H	uncertainty/flag
$K_s{}^1$	mag/upper limit	σ_{K_s}	uncertainty/flag
Scan number^1		Time/Frame number (?)	
Assoc Flag^1	(0,stellar,xgal)	Vis Magnitude	GSC/Tycho
Other Flags		V-$K_s{}^1$	for GSC/Tycho matches
J-H^1		H-$K_s{}^1$	
K_s-[12]^1		IRAS, NED pointers	

Notes: ^{1}Denotes indexed field.

2. Each scan will be broken into separate 512x1024 pixel images for each band, so they may be conveniently handled with available imaging software and hardware. Each image will be constructed from those segments of each scan in which 6 frames are coadded. Adjacent images will overlap in-scan by $0.8'$.

3. Each image will be in FITS format and provided with header information including the reference position, photometric calibration information (TBD), pixel sizes, average airmass, UTC, etc (include table of header parameters?).

4. Image pixel size will be $\sim 0.5\times$(camera pixel). An interpolation kernel will be used to estimate the brightness of each pixel in the image from the 6 overlapping frames and provide the resampling of the original under-sampled data.

5. Only artifacts due to persistence and internal reflections will be removed from the images. Radiation hits and other artifacts will be removed only to the degree to which they are corrected in the image combination algorithms.

6. Each pixel will be represented as a 2 byte integer. The scaling will be set such that the lowest value will be set to approximately 0.25-0.5 of the median noise. Sources saturated in Read$_2$-Read 1 will be saturated in the images ($K_s\sim 6.5$, H ~ 7.5, J ~ 8.5); the source catalog will have to be consulted for the correct flux.

7. Each individual overlapping frame of each image strip will have its baseline adjusted by an arbitrary constant such that the median difference between the baseline of that frame and the previous frames in the overlap zone will be zero. The first frame in each strip will not have its baseline adjusted. This technique will result in the baselines wandering during

the scan due to small warp errors in the flats and uncertainties in baseline determination, and variations in the airglow/ thermal background over the frame.

8. Focal plane distortion (such as pincushion distortion) will be corrected by 'rubber sheeting' the separate frames to a 'flat' image space. The separate bands and the overlapping frames in each band may have fractional pixel offsets with respect to each other. Each frame will have to be interpolated with an interpolation kernel to a common pixel system for each field. The 'rubber sheet' correction can be incorporated into this process, interpolating to the common pixel coordinates. The size of the kernel required will be independent of the focal plane distortion, as long as it is a small fraction of a pixel over the span of the kernel ($\sim$4 camera pixels).

9. There should be no measurable seams (< 1 DN) in flattened frames at the step separation of $\sim$50 pixels.

10. Positions derived from the images and the point source catalog will differ by less than $1''$ (1σ).

11. Magnitudes for unsaturated sources will differ in the mean by less than 0.05 mag (1σ) between the images and the point source catalog. Sources in the images will be measured using aperture photometry.

XII. PROCESSING STRATEGY AND SCHEDULING

A. Assumptions

The 2MASS project will develop the software to schedule the telescopes. Further, the project will develop on-site Q/A software to generate TBD% of the requests for rescans based on poor seeing or poor photometry without involving IPAC directly.

B. Processing

After the calibration and pointing diagnostics have been examined by the cognizant scientists/engineers, those scans that have been certified as acceptable for inclusion in the WSDB will be mapped onto a representation of the sky with $\sim 1'$ resolution. This map will contain the numbers of all scans covering that piece of sky (up to a few hundred for calibration fields).

C. Requests for Rescans

The survey coverage processor will generate requests for rescans based on all problems uncovered at earlier stages of the pipeline (calibration, PSF variations, etc.). The smallest unit for a request will be the scan. Calibration problems will require rescans of blocks of scans between calibration objects.

XIII. OVERALL PROCESSING PHILOSOPHY

A. Operations and Processing Throughput

The 2MAPPS pipeline will be designed for a minimum of human intervention except for critical quality assurance decisions requiring examination by a cognizant scientist or engineer. The pipeline and its associated hardware shall be designed so that an operator is required only to load input and output tapes into automated tape loaders at the beginning and end of the processing of a night's data. The system shall be designed to operate unattended over weekends and holidays.

The pipeline and its associated hardware shall be able to keep up with the *average* data rate from the two telescopes, assuming nominal fractions of photometric nights (30% in the north and 60% in the south). There shall be a factor of two contingency in the system design to allow for algorithm growth, computer down time, and long periods of photometric weather. There shall be another factor of two contingency to allow for reprocessing the entire data set after the completion of the survey. However, this margin may be supplied by improvements in hardware, so this capability is not required at the beginning of the project. The reprocessing period should take no longer than 1 year, taking into account reasonable expectations for improvements in computing power over the duration of the project. Combining these factors leads to the requirement that the system be able to process 20 hours of data (10 from each of two telescopes) in roughly 12 hours of clock time.

B. Portability

There is a long term goal to make the 2MAPPS pipeline easily portable to other platforms with a minimum of recoding. To this end it is a requirement that the processing pipeline software be written in ANSI standard Fortran77 or ANSI standard C. No assembly language software will be permitted in the pipeline without explicit agreement of the Project Manager. Machine-dependent or non-portable code, if required, shall be carefully isolated in a few, well identified subroutines. The system executive shall also be written in such a way as to facilitate porting to other systems.

C. Software Change Control

IPAC will exercise strict change control over the 2MAPPS pipeline software, including centralized delivery of source code, centralized make files, and automated regression testing. Analysis and ancillary support software will not fall under strict control.

XIV. ARCHIVING

Copies of the raw data will be maintained at IPAC and UMASS.

The 2MASS data volume is so large ($\geq$ 10 Tbyte) that the only access that IPAC can afford is operator-mounted tapes. The option of robotic access to the raw data with high density optical media will be re-examined periodically, but is not required for pipeline processing of the data.

The 2MASS images will similarly be archived at IPAC on 8-mm tapes with a backup copy to be stored at UMASS. The option of robotic access to the raw data with high density optical media will be re-examined periodically.

Images of selected areas, e.g., the calibration scans, will be archived separately for easy access for technical and scientific analysis.

IPAC will investigate lossy compression techniques for the image data that may reduce the data volume by factors of 10-100, at the cost of some photometric and positional accuracy.