MIT Libraries | DSpace@MIT

MIT Open Access Articles

ANDROMEDA (M31) OPTICAL AND INFRARED DISK SURVEY. I. INSIGHTS IN WIDE-FIELD NEAR-IR SURFACE PHOTOMETRY

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Sick, Jonathan, Stephane Courteau, Jean-Charles Cuillandre, Michael McDonald, Roelof de Jong, and R. Brent Tully. "ANDROMEDA (M31) OPTICAL AND INFRARED DISK SURVEY. I. INSIGHTS IN WIDE-FIELD NEAR-IR SURFACE PHOTOMETRY." The Astronomical Journal 147, no. 5 (April 10, 2014): 109. © 2014 The American Astronomical Society

As Published: http://dx.doi.org/10.1088/0004-6256/147/5/109

Publisher: IOP Publishing

Persistent URL: http://hdl.handle.net/1721.1/92933

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Massachusetts Institute of Technology

The Astronomical Journal, 147:109 (19pp), 2014 May doi:10.1088/0004-6256/147/5/109
© 2014. The American Astronomical Society. All rights reserved. Printed in the U.S.A.

ANDROMEDA (M31) OPTICAL AND INFRARED DISK SURVEY. I. INSIGHTS IN WIDE-FIELD NEAR-IR SURFACE PHOTOMETRY

Jonathan Sick[1], Stéphane Courteau[1], Jean-Charles Cuillandre[2], Michael McDonald[3], Roelof de Jong[4], and R. Brent Tully[5]
[1] Department of Physics, Engineering Physics & Astronomy, Queen's University, Kingston, Ontario, K7L 3N6, Canada; jsick@astro.queensu.ca
[2] Canada–France–Hawaii Telescope Corp., Kamuela, HI 96743, USA
[3] Kavli Institute for Astrophysics and Space Research, MIT, Cambridge, MA 02139, USA
[4] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany
[5] Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI, USA
Received 2013 March 26; accepted 2014 January 21; published 2014 April 10

ABSTRACT

We present wide-field near-infrared J and K_s images of the Andromeda Galaxy (M31) taken with WIRCam at the Canada–France–Hawaii Telescope as part of the Andromeda Optical and Infrared Disk Survey. This data set allows simultaneous observations of resolved stars and near-infrared (NIR) surface brightness across M31's *entire* bulge and disk (within $R = 22$ kpc), permitting a direct test of the stellar composition of near-infrared light in a nearby galaxy. Here we develop NIR observation and reduction methods to recover a uniform surface brightness map across the $3^\circ \times 1^\circ$ disk of M31 with 27 WIRCam fields. Two sky-target nodding strategies are tested, and we find that strictly minimizing sky sampling latency cannot improve background subtraction accuracy to better than 2% of the background level due to spatio-temporal variations in the NIR skyglow. We fully describe our WIRCam reduction pipeline and advocate using flats built from night-sky images over a single night, rather than dome flats that do not capture the WIRCam illumination field. Contamination from scattered light and thermal background in sky flats has a negligible effect on the surface brightness shape compared to the stochastic differences in background shape between sky and galaxy disk fields, which are ∼0.3% of the background level. The most dramatic calibration step is the introduction of scalar sky offsets to each image that optimizes surface brightness continuity. Sky offsets reduce the mean surface brightness difference between observation blocks from 1% to <0.1% of the background level, though the absolute background level remains statistically uncertain to 0.15% of the background level. We present our WIRCam reduction pipeline and performance analysis to give specific recommendations for the improvement of NIR wide-field imaging methods.

Key words: methods: observational – techniques: photometric

Online-only material: color figures

1. INTRODUCTION

Thanks largely to its proximity, kinship with the Milky Way, and being an ideal foil for modern galaxy formation models, the Andromeda Galaxy (M31) has been the focus of numerous investigations of galaxy structure (Ibata et al. 2005; Irwin et al. 2005; McConnachie et al. 2009; Courteau et al. 2011) and stellar populations (Williams 2002; Worthey et al. 2005; Saglia et al. 2010). The shapes, ages, kinematics, and relative fraction of galaxy components (bulge, disk, halo) in large spiral galaxies like our own reveal precious information about their formation, accretion, and merging histories (see the review of Kormendy & Kennicutt 2004).

Unfortunately, the fundamental task of disentangling galaxy components—which typically involves light profile decompositions, color gradients, and mass modeling—remains non-trivial. Stellar population studies of spiral galaxies are thwarted by a threefold degeneracy between stellar age (A), metallicity (Z), and interstellar medium dust that can only be lifted by combining, at the very least, optical and infrared images with realistic dust models (de Jong 1996; MacArthur et al. 2004; Pforr et al. 2012). Likewise, mass models of spiral galaxies suffer a degeneracy between the stellar M/L and dark halo parameters that requires, in addition to an extended rotation curve, deep and accurate multi-band imaging to uniquely constrain the stellar M/L ratio (Dutton et al. 2005; Courteau et al. 2014).

Compounding these challenges are fundamental uncertainties in modern stellar population synthesis, particularly uncertainties in the interpretation of near-infrared (NIR) light. For instance, optical–NIR spectral energy distributions (SEDs) yield unreliable population synthesis fits compared to optical-only SED fits. This failure is largely attributable to inadequate stellar population synthesis recipes for NIR bands and naive parameterization of star formation histories (SFHs; Taylor et al. 2011; Courteau et al. 2014).

First, SED fitting often relies on simplistic SFH parameterizations. Because NIR colors lift age–metallicity–dust degeneracies, modeling of NIR bands may require additional sophistication, namely, composite star formation and metal enrichment histories. The appropriate form of SFH models cannot be constrained from the integrated light of galaxies alone (as is typically attempted); resolved color–magnitude diagrams (CMDs) are both more effective and, in fact, essential for deriving non-parametric stellar population histories.

Second, asymptotic giant branch (AGB) stars from intermediate-aged stellar populations heavily influence the NIR light (Maraston 1998). Modeling AGB stars is most challenging due to their complex dredge-up cycles that change surface chemistry and temperature (the M- to C-type transition) and circumstellar winds that further perturb an AGB star's location in the CMD. A proper calibration of NIR stellar population synthesis models (e.g., Maraston 2005; Marigo et al. 2008; Charlot

& Bruzual, in preparation) may yield a 30%–50% improvement in the estimation of stellar masses and ages of high-redshift systems (e.g., Maraston et al. 2006; Bruzual 2007; Conroy & Gunn 2010; Conroy 2013).

Our remedy for understanding the structure and NIR stellar populations of M31 is to survey the entire bulge and disk ($R \leqslant 22$ kpc) in both resolved and integrated stellar light at J and K_s wavelengths. In doing so, we can directly relate an NIR stellar population's decomposition in the color–magnitude plane to the panchromatic SED of M31. Though such a calibration could be made with other galaxies, M31 is unique in its proximity so that even ground-based instrumentation can resolve its bright stellar population. For reference, $1'' = 3.7$pc across the disk of M31 (we adopt $D_{\rm M31} = 785$ kpc; McConnachie et al. 2005).

A wealth of photometric data exist for M31; however, none provide simultaneous resolved and integrated maps of the NIR light. The best resource for resolved stellar populations across the entire disk of M31, to date, is the Local Group Galaxy Survey (LGGS; Williams 2003; Massey et al. 2006). This survey, although covering the *UBVRI* wavelengths, does not extend to the *JHK* NIR wavelengths most contentious for stellar population models. Advancing our view of M31 to NIR wavelengths, Beaton et al. (2007) assembled a 2.°8 JHK_s mosaic of M31 with the Two Micron All Sky Survey (2MASS) 6X program. Those observations were also used by Athanassoula & Beaton (2006) as evidence for a bar embedded in a classical bulge. Beyond the bulge, the 2MASS 6X images have limited utility; background level uncertainties restrict their use to inferring structural and photometric properties of the disk. Further, the pixel scale of $1''$ and integration depth of 46.8 s prevent point-source measurements of M31 stars in the 2MASS 6X images. As a result, the then state-of-the-art NIR view of M31 is the 3.6 μm *Spitzer*/IRAC map of Barmby et al. (2006). While *Spitzer* reduces uncertainties in background estimation, the pixel scale of 0.″86 still prevents point-source measurements of individual stars in the M31 disk.

Decompositions of M31's stellar populations have been made with high-resolution *Hubble Space Telescope* (*HST*) observations (Brown et al. 2003, 2006, 2008). These authors detected an intermediate-age population in the inner halo of M31, and even disentangled debris associated with the Giant Stream around M31 in six fields sampling the outer disk, halo, and Giant Stream around M31 (Brown et al. 2009). Similarly, though in the near-infrared, Olsen et al. (2006) derived SFHs within $22.''5 \times 22.''5$ fields in M31's inner disk and bulge using ground-based adaptive optics observations with the Gemini ALTAIR/NIRI instrument. However, these pencil-beam surveys cannot be construed as representative of the entire Andromeda Galaxy. Thus, the boldest step forward in understanding M31's stellar populations is coming from the Panchromatic Hubble Andromeda Treasury Survey (PHAT; Dalcanton et al. 2012). PHAT provides wide-field coverage from M31's center to just beyond the 10 kpc star-forming ring, a panchromatic view of stellar populations from 3000 Å to 17000 Å, and resolution of stars in very crowded environments such as the bulge. Note that PHAT only covers a single quadrant of M31, making any global conclusions about M31's stellar populations incomplete. Further, the wavelength coverage of *HST*/WF3 falls short of the 2.2 μm K band, making empirical calibration of the common NIR bands used by wide-field NIR surveys (*JHK*) impossible. Thus, there is good cause, even in the era of PHAT, to revisit M31 with a ground-based survey that covers the entire disk of M31 at NIR wavelengths, while using the best natural seeing in the Northern Hemisphere (on Mauna Kea) to resolve stars even more effectively than the previous ground-based survey of M31's disk (LGGS).

We present the first installment of such a survey, the Andromeda Optical and Infrared Disk Survey (ANDROIDS), in this paper. ANDROIDS uses the WIRCam instrument (Puget et al. 2004) on the Canada–France–Hawaii Telescope (CFHT), which is among the first generation of wide-field ground-based NIR detector arrays, covering a $21.'5 \times 21.'5$ field of view. Indeed, the advent of detectors such as WIRCam makes such a wide-field, high-resolution survey of an object as vast as M31 possible. The excellent natural seeing (0.″65) on Mauna Kea is sufficient for resolving giant branch stars throughout the disk of M31.

Recovering the true NIR surface brightness map of M31 is, however, technically challenging. The NIR background (from both atmospheric fluorescence and instrumental thermal emission) is $\sim 10^3\times$ brighter than the NIR surface brightness of M31 at $R = 20$ kpc, demanding exceptional background characterization. Whereas most NIR galaxy surveys can measure the instantaneous background from blank sky pixels surrounding the galaxy on a detector array, M31's extended size requires physically nodding the telescope away from the galaxy by 1°–3° to sample blank sky (called sky-target, or ST, nodding). That we can never observe the instantaneous background on the disk of M31, but rather sample the sky at both a different location and time, introduces additional complications. Adams & Skrutskie (1996) clearly showed, with $9^\circ \times 9^\circ$ movies of the sky, that NIR sky emission has coherent spatial structure that moves across the sky over those scales, akin to a cirrus cloud system. This assures that background sampled from a sky field *will not* correspond directly to the sky affecting disk observations.

Besides background level, another concern is the accuracy of the surface brightness shape across individual WIRCam fields of view. Spatial structures in the NIR sky can leave residual shapes in background-subtracted disk images that ultimately affect our ability to produce a seamless NIR mosaic of M31. Vaduvescu & McCall (2004) also found that detector systems themselves, in their case the (now decommissioned) CFHT-IR camera, can add a time-varying background signal whose strength may be comparable to the NIR surface brightness of the outer M31 disk.

Because such a large mosaic has never before been assembled in an ST nodding WIRCam program, we focus this contribution on engineering the best practices for this type of observing. This includes: finding the optimal ST nodding cadence, defining the appropriate data reduction procedures for a WIRCam surface brightness reduction, and finally presenting an analysis of the surface brightness accuracy in wide-field WIRCam mosaics.

Section 2 describes the novel observational strategies used to reduce background subtraction uncertainties. Section 3 presents the image reduction pipeline, followed by night-sky flat fielding and median background subtraction in Section 4, and zero-point calibration practices in Section 5. The accuracy of our WIRCam image calibrations is analyzed in Section 6. In Section 7 we present our method for recovering the galaxy surface brightness by minimizing image-to-image differences across the mosaic, while in Section 8 we analyze the results of this algorithm. We estimate the systematic uncertainties in our mosaic solution in Section 9, where we also compare our technique to the Montage package (Berriman et al. 2008) and the *Spitzer*/IRAC mosaics. Finally, in Section 10 we summarize the uncertainty of NIR background subtraction on the scale of M31 and outline our ideal observation and reduction method. In a future installment we will use resolved star counts to establish the NIR surface brightness of the M31 disk beyond $R > 15$ kpc.

Table 1
Summary of WIRCam Observing Programs

Semester	Band	$N_{\rm disk}$	ST Nods	$T_{\rm int}$/field (minutes)	$T_{\rm exp}$ (s)	Eff. (%)	$\mu_{\rm bkg}$ (mag arcsec^{-2})	PSF FWHM (arcsec) 25th	50th	75th
2007B	J	27	$[S^3T^8]^2S^3$	12.5	47	49	(15.4, 16.7)	0.68	0.75	0.84
	K_s		$[S^5T^{13}]^2S^5$	10.8	25	42	(13.4, 14.2)	0.60	0.65	0.73
2009B	J	12	$[ST^2]^{20}S$	13.3	20	26	(15.0, 16.5)	0.61	0.69	0.83
	K_s						(13.4, 14.3)	0.60	0.66	0.76

Notes. $N_{\rm disk}$ is the number of WIRCam fields covering the M31 disk in each semester (see Figure 1). ST nods with superscripts denote the number of times an observation is repeated for a given field. $T_{\rm int}$ is the total integration time per disk field, while $T_{\rm exp}$ is the integration time per WIRCam exposure. Eff. is the observing efficiency, or percentage of time in a program allocated to integrating the disk of M31, compared to nodding, read out, and sky overheads. $\mu_{\rm bkg}$ gives the range (min–max) of background surface brightnesses seen in each band. Point-spread function (PSF) reports the distribution seeing as measured from the FWHM of stellar PSFs in the uncrowded sky images.

2. OBSERVATIONS

M31 was observed in the NIR using the WIRCam instrument, mounted to the 3.6 m CFHT, at the summit of Mauna Kea in Hawaii over multiple runs between 2007 and 2009. Observations were carried out in the NIR J ($\lambda_0 \sim 1.2$ μm) and K_s ($\lambda_0 \sim 2.2$ μm) bands.

WIRCam is an array of four HgCdTe HAWAII-RG2 detectors (Puget et al. 2004). Each detector comprises 2048×2048 pixels, with a scale of $0\farcs3$ pixel^{-1}, which critically samples CFHT's typical seeing of $0\farcs65$. WIRCam's detectors are arranged in a 2×2 grid with $45''$gaps, so that the entire instrument covers $21\farcm5 \times 21\farcm5$ of sky. It is truly the advent of NIR focal plane arrays, like WIRCam, that has enabled wide-field NIR studies of M31.

The ANDROIDS WIRCam survey is designed to simultaneously resolve stars and recover the integrated surface brightness of the entire M31 disk. As discussed in Section 1, NIR observations require frequent monitoring of the background. Vaduvescu & McCall (2004) found, for example, that the NIR background intensity can vary by 0.5% per minute; yet the low surface brightness of M31's NIR disk at $R = 20$ kpc requires the background level to be constrained to approximately 0.01% (equivalently, ~0.01 mag). With a $190' \times 60'$ optical disk, M31 is much larger than the WIRCam fields of view, and monitoring of the background zero point is only possible by periodically pointing the telescope away from M31 toward blank sky through ST nodding. The fundamental compromise of ST nodding observation programs is to balance the cadence of sky sampling with the efficiency of observing the target itself. Although studies such as Vaduvescu & McCall (2004), and references therein, provide good guidelines for NIR background behavior, no program has attempted to construct an NIR surface brightness mosaic covering an area as large and faint as M31's disk.

We now have the opportunity to experiment with different ST nodding strategies since observations were taken over the 2007B and 2009B semesters. An objective of this study is to determine how observational design can improve the construction of a wide-field NIR mosaic by comparing the performance of two pre-defined observing strategies.

2.1. 2007B Semester

The initial survey was carried out in the 2007B semester by the CFHT's queue service observing under photometric conditions. Here M31 is covered with 27 contiguous WIRCam fields out to the optical radius where $\mu_V = 23$ mag arcsec^{-2} at $R = 20$ kpc. The fields are arranged with at least $1'$ overlap in declination, and approximately $5'$ overlap in right ascension. The field configuration is shown in Figure 1.

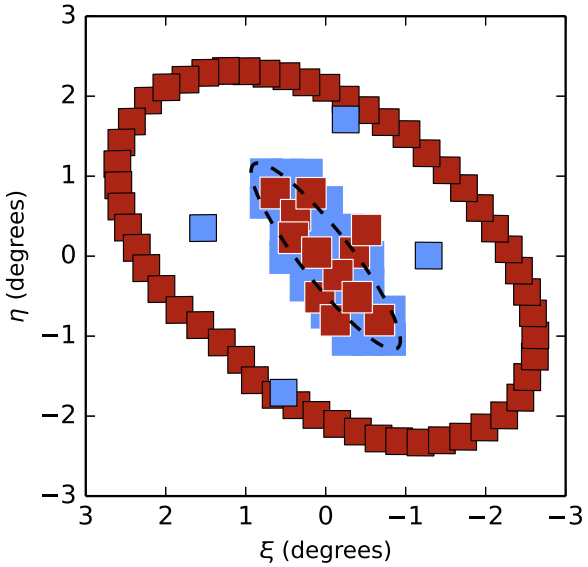

Figure 1. ANDROIDS WIRCam field positions on M31. Central blue tiles are the 27 disk fields observed in 2007B, surrounded by four sky fields. Dark red tiles at center are the 12 disk fields observed in 2009B. The outer ring of 53 fields is the 2009B sky sampling ring. The dashed ellipse marks the M31 disk at $R = 20$ kpc along the major axis. Coordinates are centered on the nucleus of M31 with north up, and east to the left.

(A color version of this figure is available in the online journal.)

As shown in Table 1, each field was integrated for 16×47 s = 12.5 minutes in J and 26×25 s = 10.8 minutes in K_s. These integrations are sufficiently deep for resolved stellar photometry to reach at least 1 mag below the tip of the red giant branch (RGB), a crucial requirement for decomposing the contributions of red giant and AGB stars to the NIR light.

The 2007B ST nodding strategy was motivated by a canonical understanding of NIR background behavior, since ST nodding background subtraction had never been attempted on this scale before. The NIR background intensity can be expected to change by 5% in 10 minutes (Adams & Skrutskie 1996; Vaduvescu & McCall 2004). Since the background itself is $\sim 10^3\times$ brighter than the outer disk of M31 in the NIR, a 5% uncertainty in the background would be fatal to our objective of recovering M31's extended NIR surface brightness. To constrain background brightness within 1%, we ensured that a sky sample would be no

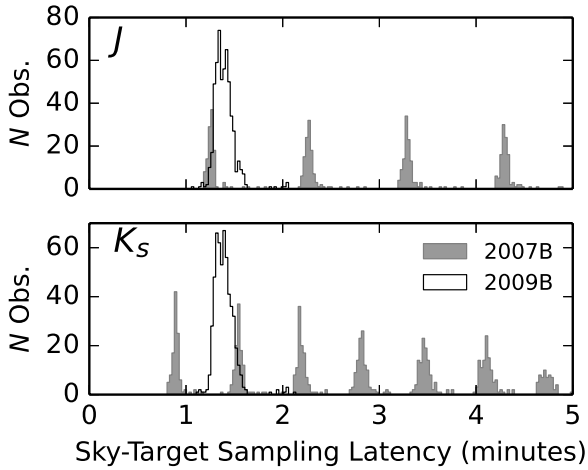

Figure 2. Time latency between target observations and sky field sampling in the 2007B and 2009B WIRCam observing runs. The 2009B program was designed to ensure that no disk sample would be removed by more than 1.5 minutes from a sky sample by using an STTS nodding pattern.

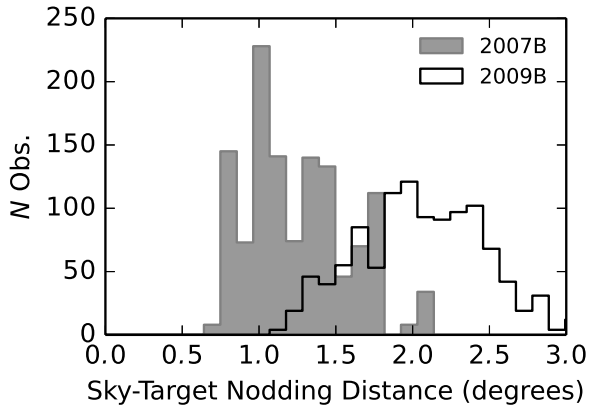

Figure 3. Distance between sky and target observations in the 2007B and 2009B WIRCam observing runs. The larger nodding distance of 2009B is a consequence of sky ring sampling. The maximum nodding distance across the sky ring was purposely set to ∼3° to avoid excessive time overheads (see Figure 1). As such, a given disk field only samples roughly half of the full sky ring.

more than 5 minutes removed from an M31 target image. Given the respective exposure times (chosen so as not to saturate with the background flux), this implied a sky (*S*)–target (*T*) observing sequence of $S^3T^8S^3$ in *J* and $S^5T^{13}S^5$ in K_s to minimize telescope repositioning overhead.[6] Four sky fields were chosen (Figure 1), and each disk field was associated with a single sky field.

2.2. 2009B Semester

Analysis of the 2007B data revealed that the adopted ST nodding strategy was not sufficient for recovering the M31 surface brightnesses due to uncertainties in the background. Repeatedly sampling one of only four sky fields also proved not ideal. This motivated the 2009B observing campaign.

Rather than replicate the 28 field footprint of the 2007B campaign, we observed 12 new fields in 2009B (see red boxes in Figure 1) that overlap each other and all of the 2007B footprints, to form a network of well-subtracted fields. Thus, the 2009B observations augment and calibrate the 2007B NIR mapping.

To improve background subtraction fidelity, we recognized challenges not fully appreciated in the 2007B survey design. Not only does the background change rapidly in time, but it possesses a significant spatial structure on the scale of WIRCam fields and larger. This has two ramifications: the background level sampled at a sky field *will not* necessarily reflect the background present at the disk, and the background in each WIRCam frame has a two-dimensional shape, not simply a scalar level.

This resulted in three principal changes to the observing strategy. First, we chose to minimize latency between sky and target observations with an ST^2S pattern. That is, each target observation was directly paired with a sky observation taken within 1.5 minutes (Figure 2).

Second, we also increased the number of repetitions so that each field is observed 40 times in each band in an $[ST^2]^{20}S$ pattern. This repetition enables averaging over spatial sky background structures on the scale of WIRCam fields.

Finally, we employ a pseudo-randomized ST nodding pattern where no sky field is used repeatedly for a disk field. In order to maintain rapid telescope nods, only northern sky fields serviced the northern disk, and similarly for the southern fields; the maximum offset on the sky was 3° (see Figure 3). This non-repetitive sampling of sky fields yielded two possible advantages: (1) when a median background image is constructed, many *background shapes* are combined, possibly yielding an intrinsically flatter background image (see Section 4.1.2); and (2) if there is a coherent structure in the NIR background, sampling sky fields degrees apart in rapid succession should average out these systematic biases in estimating the background level *on the galaxy disk*. Given these observations, we now consider how to properly construct a wide-field NIR surface brightness mosaic of M31.

3. IMAGE PREPARATION

While CFHT distributes calibrated WIRCam data products, we have chosen to replace much of their data reduction recipes with our own to optimize and explore the limitations of wide-field NIR surface brightness maps. An overview of the pipeline is shown in Figure 4; the principal steps are (1) astrometry, (2) source masking, (3) night-sky flat fielding, (4) zero-point estimation against 2MASS sources, (5) median sky frame construction, (6) image calibration with zero points and median sky subtraction, and (7) background optimization and mosaic production in three hierarchical steps.

3.1. Choice of Starting Point

CFHT offers WIRCam data in two degrees of preprocessing with the *'I'iwi* pipeline: an image that has been corrected for nonlinearity, dark subtracted, and flat fielded (`*s.fits`); and an image that has been background subtracted, in addition to all the previous treatments (`*p.fits`). In order to implement our own calibration strategy, our mosaics stem from `*s.fits` products (though we note that `*p.fits` products are still used for astrometry and source masking; see below). Nonetheless, two *'I'iwi* processing stages included in `*s.fits` products must be handled carefully.

Cross-talk correction. WIRCam integrations prior to 2008 March (which includes the 2007B data set, not the 2009B data) suffered from electronic cross-talk within the detector. This cross-talk is manifested in repeating rings above and

[6] Superscripts here denote the number of times an observation is repeated in sequence for a given target disk field.

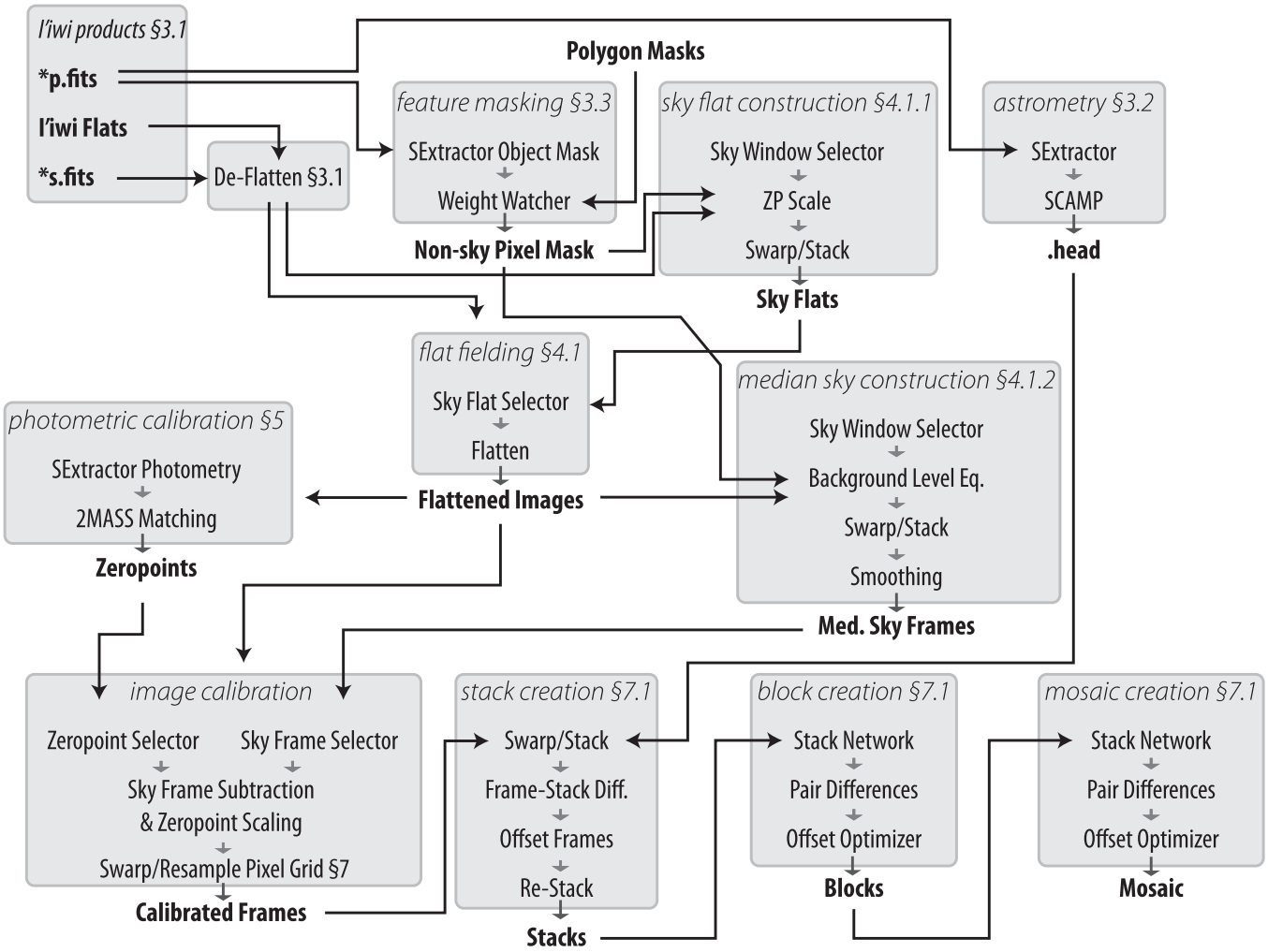

Figure 4. Flowchart representation of the ANDROIDS WIRCam pipeline, from receipt of CFHT *'I'iwi* data products to rendering of M31 mosaics.

below saturated stars.[7] By default, the *'I'iwi* pipeline removes this cross-talk by subtracting a median of the 32 amplifier slices. Unfortunately, this algorithm fails in cases where the background has a surface brightness gradient (such as on the disk of M31) and produces a brightness gradient that is stronger than the galaxy surface brightness itself. Loic Albert (then at CFHT) kindly re-processed our 2007B data set with the cross-talk correction omitted.

Flat fielding. A peculiarity of `*s.fits` images is that even though they are flat fielded using dome flats by CFHT, those products still exhibit strong non-uniformity, dust artifacts, and surface defects. We note that CFHT produces dome flats (for each queue run) from median stacks of 15 images taken under a tungsten lamp, subtracted from images of the same integration time taken with the lamp off. This procedure should remove the additive thermal background from the flat, ensuring that the flat field is a purely multiplicative calibration. Despite this, the presence of dust artifacts betrays the fact that dome flats do not reproduce the same illumination pattern as sky photons. Similarly, the presence of surface defects in dome-calibrated images could be caused by disparities in both the optical path and the color of the tungsten lamp versus the night-sky background. In this work, we find that WIRCam images can be adequately flat fielded using night-sky flats. We give a visual demonstration of the superiority of night-sky flat fielding in Figure 5: surface defects left by dome flat fielding are removed with night-sky flat fielding.

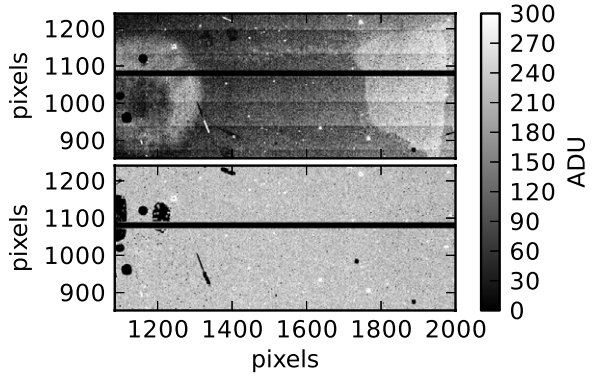

Figure 5. Comparison of a WIRCam frame cutout processed with dome flats by the *'I'iwi* pipeline (top), and with sky flats (bottom). Both images are shown in linear counts with identical level ranges. No median background subtraction has been applied. Dome flats leave WIRCam images with dust artifacts (left) and detector surface defects (right). Furthermore, the 64 pixel high horizontal amplifier bands are clearly visible. Simply using sky flats eliminates these artifacts.

One interesting feature of `*s.fits` images is the appearance of horizontal banding corresponding to the 32 amplifiers that

[7] See http://cfht.hawaii.edu/Instruments/Imaging/WIRCam/WIRCamCrosstalks.html.

service independent horizontal bands of each WIRCam detector (seen in Figure 5). It is odd that a dome flat failed to calibrate such electronic structures in a detector, and one might expect that such banding should be calibrated with an additive correction. Indeed, this banding is absent from fully processed *'I'iwi* images due to median sky frame *subtraction*. However, we maintain that flat fielding is the correct treatment for these structures since they appear to be proportional to the background throughout the night (which can vary by 10% during a night), yet are still corrected with a single night-sky flat.

Our ANDROIDS pipeline thus begins with `*s.fits` data that have been *uncorrected* for dome flat fielding. That is, we multiply the `*s.fits` image with its associated dome flat.[8] The result is an image that retains *'I'iwi*'s prescription for dark subtraction, bad pixel masking, and nonlinearity correction, ready for our own sky flat fielding (to be described in Section 4).

3.2. Astrometry

Early in the pipeline we build a unified astrometric frame for our image set using *SCAMP* (Bertin 2006). *SCAMP* matches stars in *Source Extractor* (Bertin & Arnouts 1996) catalogs of each WIRCam frame both internally (to $\sigma_{\rm int} = 0\farcs10$) and against the 2MASS Point Source Catalog (PSC; Skrutskie et al. 2006) to a precision of $\sigma_{\rm ref} = 0\farcs15$. By processing all 4286 frames in the ANDROIDS/WIRCam survey simultaneously, *SCAMP* allows an accurate and internally consistent coordinate frame for our mosaic. *SCAMP* handles this data volume gracefully provided we cull the input star catalogs for stars with signal-to-noise ratio$>$ 100, and by using the `SAME_CRVAL` astrometry assumption that the WIRCam focal plane geometry is stable. Also note that we build our *Source Extractor* catalogs using the fully processed `*p.fits` *'I'iwi* images since those are adequate for source detection and astrometry. While *SCAMP* is capable of also fitting a photometric solution for each frame, we choose to establish photometric zero points later in our pipeline using a combination of background flux observed across the detector array and bootstrapping against 2MASS sources observed in uncrowded 2MASS images (see Sections 4.1.1 and 5).

3.3. Non-sky Pixel Masking

A second preliminary pipeline stage is source masking. For each sky image we build masks that yield only blank pixels to aid with background level estimation, sky flat construction (Section 4), and median sky frame construction (Section 4.1.2). These masks are built by a combination of *Source Extractor* object maps (detected in `*p.fits` images) and hand-drawn polygon regions that cover the diffraction spikes and halos of very bright foreground stars. These masks, along with the *'I'iwi* bad pixel mask, are combined with *WeightWatcher* (Marmo & Bertin 2008).

4. SKY FLAT FIELDING AND MEDIAN SKY SUBTRACTION

As we mentioned previously in Section 3, dome flats fail to properly calibrate dust, amplifier gain, and surface defects in WIRCam data (see Figure 5). Sky flats are an appropriate alternative, both because of the abundant background photons (any NIR imaging program can use its own images to build sky flats) and because sky flats directly match the illumination path of observations. Sky flats also have the advantage of being contemporaneous with observations: if the WIRCam flat field is variable, then sky flats can be built to track such variability. This is a distinct advantage over dome flats, which CFHT builds at the beginning of every WIRCam queue run, or even twilight flats that can only be built once per night.

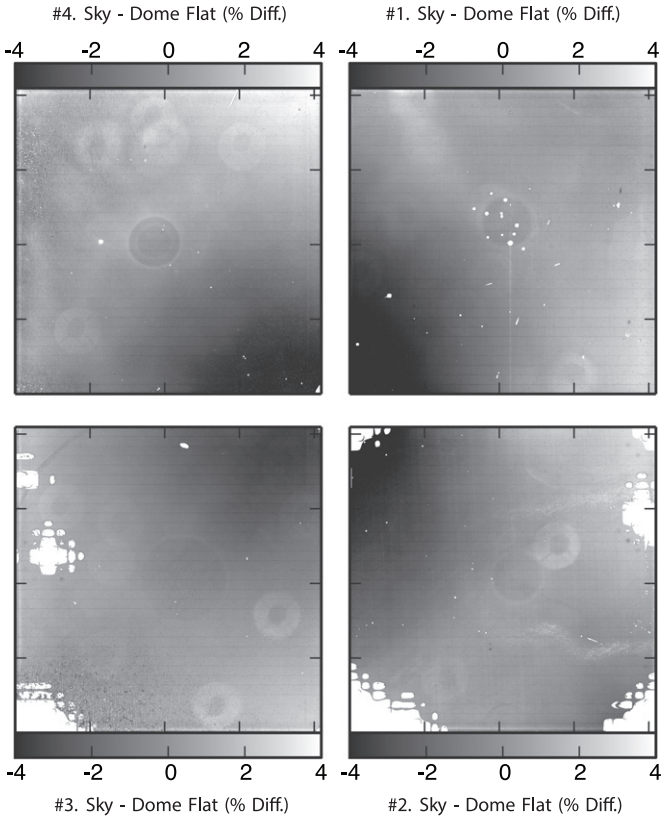

Figure 6. Ratio of corresponding K_s band `NIGHT` sky flat and dome sky flats. This ratio provides an approximate upper limit on the systematic uncertainty of flat fielding with WIRCam. Note that the `NIGHT` sky flats have been renormalized to remove the chip-to-chip zero-point correction described in Section 4.1.1.

Despite these advantages, sky flats are built on the assumption that all background illumination in a night-sky image is proportional to the flat-field function. Several contaminants prevent this from being true: thermal emission from the detector or telescope structures can add a significant background in the K_s band, and scattered light (e.g., off the camera's cold pupil stop) further perturbs the proportionality of flat-field images. Ideally one would subtract these contaminants from images before constructing sky flats. Then science images could be flat fielded, and additive contaminants in science images would be automatically removed in subsequence median background subtraction. Note that both dome flats and twilight flat fields can distinguish additive contaminants from the multiplicative flat-field function. Dome flats are built from the differences of images taken with the lamp on and off, directly removing any thermal component from the flat field. Twilight flats can also treat additive contamination by capturing images at different levels of sky illumination so that linear fits to each pixel allow any additive bias to be removed. In the case of sky flat fielding, however, we *cannot* disentangle additive from multiplicative processes in images so that the sky flat field and median background subtraction steps presented here are not in fact separable and independent operations. Figure 6 shows a ratio of sky flats (using the `NIGHT` prescription; see below) and dome flat. This ratio effectively

[8] Dome and twilight flats are made available by CFHT, http://limu.cfht.hawaii.edu:80/detrend/wircam/.

sets an approximate upper limit on the systematic accuracy of a WIRCam flat field. Thus, we can expect our flats to be correct within a few percent. To proceed, we must accept that our sky flats are not purely proportional calibrations and instead are a first step in our combined flux calibration and background subtraction pipeline. In Section 6 we analyze the performance of our night-sky flat fielding and median background subtraction procedure to find that frame-by-frame surface brightness shape errors are dominated by rapid variations in the background itself.

A second assumption built into sky flat construction is that skyglow is uniform across the detector. Wide-field images of the NIR night sky show that skyglow has rich spatial and temporal variations. However, by marginalizing over a large number of sky images, any illumination bias in the sky can be mitigated. The 2009B observing program even took this marginalization process further by sampling pseudo-random sites on the sky while building sky flats. One degree of control that can be exerted over sky flat construction is the time window that sky images are drawn from. Using a long window, such as the full length of a queue observing run, ensures that the intrinsic WIRCam flat-field function is stable over several days, while producing the statistically flattest residual sky illumination pattern. Shorter windows make the opposite assertion that the WIRCam flat field is unstable, and that any bias in sky shapes can be tolerated.

We investigate three sky flat designs in this study, labeled QRUN, NIGHT, and FW100K. QRUN flats are built from all of the sky integrations taken during a queue run and through a given filter. For the ANDROIDS program, 25–637 (typically ∼140) sky images, obtained over a sequence of ⩽10 days, are composed into a QRUN flat. NIGHT flats are made from all of the sky integrations taken during a single night, through a given filter. Because observations are observed in queue service observing mode, the ensemble of sky images typically sample 0.5–3 hr of a night. Last, we introduce *real-time* FW100K sky flats that are rapidly updated throughout the night in case the WIRCam illumination function and gain structure are unstable. FW100K flats are designed such that the pool of sky images reaches cumulative background levels of at least 100,000 ADU, or that the time span from first to last sky integration is no longer than 2 hr. Given the 07B *J*-band ST nodding pattern, 15 sky integrations are accumulated in 50 minute windows, whereas the more frequent nodding in the 09B campaign shortened this window to 20 minutes (though as long as 50–90 minutes in dark sky conditions). The brighter K_s sky calls for just 7–13 integrations in 07B, or 10–20 integrations in the 09B campaign. This number of K_s sky samples was accumulated within 10–30 minutes in 07B, or 10–70 minutes in 09B.

4.1. Implementation of Sky Flat Fielding and Median Background Subtraction

We now describe the technical details of sky flat fielding and median background subtraction steps. Recall from Figure 4 that the inputs of flat-field construction are "de-flattened" images that retain the linearity and dark-current subtraction of the *'I'iwi* pipeline.

4.1.1. Sky Flat Construction

According to the type of sky flat being constructed, QRUN, NIGHT, and FW100K, ensembles of sky images are formed. Given an ensemble of sky integrations, our next task is to scale the intensity of each image according to three requirements: (1) each image frame in the median stack is at the same level, (2) each WIRCam detector has a unified zero point, and (3) the sky flat across the whole array is flux normalized.[9] This scaling is determined by the median pixel level measured on each detector for each sky integration—let us denote these median levels as $\alpha_{i,j}$ for the ith sky image's level in detector j ($j \in \{1, 2, 3, 4\}$). To avoid bias in the background estimate, we mask any pixels that do not sample blank sky (see Section 3.3).

From the ensemble of images produced by an individual WIRCam detector, we compute the median background level: $\beta_j = \mathrm{median}(\alpha_{1,j}, \alpha_{2,j} \ldots \alpha_{n,j})$. Further, we also compute S, the median of all median detector levels: $S = \mathrm{median}(\beta_1, \beta_2, \beta_3, \beta_4)$. Then each sky image is scaled by the factor $f_{ij} = \beta_j/(\alpha_{ij} S)$. Note that the factor α_{ij}^{-1} normalizes each image to the same level for stacking, while the ratio β_j/S adjusts the level of each detector according to detector-to-detector zero-point offsets.

The flat itself is built by median combination. Median combination of a stack of hundreds of 2048 × 2048 pixel images, each with a weightmap masking astronomical sources, is computationally intensive. A convenient solution is to use *Swarp* (an image-mosaicking software package, Bertin et al. 2002) in a mode that combines images pixel-to-pixel. Once the sky flat is built, it is divided from the appropriate science images to produce a flat-fielded data set.

4.1.2. Median Background Subtraction

Since M31 is much larger than individual WIRCam fields, background is subtracted (to first order) using the background levels found in contemporary sky images. Section 2 described the ST nodding sequences chosen for the 2007B and 2009B observing campaigns. Although a scalar background level can be estimated from a sky image and subtracted from the paired target images, it is common to construct a median background image and subtract this from target images.

Independent median background images for each WIRCam detector are produced by choosing a sky image (the primary sky image) and four other sky images taken at adjacent times. Across each image, the median background intensity is recorded. A Source Extractor object mask, as used in Section 4.1 for flat fielding, removes bias from astrophysical sources. Each sky image is additively scaled to a common intensity level to compensate for background level variations. As described in Section 4.1, *Swarp* is used to median-combine the sky images with non-sky pixel masks. Since the background has only low-frequency spatial information, these median background images are smoothed with a Gaussian kernel (note this is quite different from the function of median background images applied to dome-flat processed WIRCam data, where median background subtraction also removed pixel-to-pixel artifacts). This median background image is then additively scaled back to the original level of the primary sky image. Finally, to background subtract a science image, we apply the concurrent median background image.

5. PHOTOMETRIC CALIBRATION

Our flat-field procedure necessitates a revision of photometric zero points. Since our program is observed in short (∼1 hr) blocks in CFHT's queue service observing, we do not have the necessary airmass baseline to solve for nightly zero-point

[9] It is also acceptable to establish chip-to-chip zero-point offsets using differential 2MASS photometry, rather than from background surface brightness. In Section 6.2 we establish the equivalence of the two methods.

and atmospheric extinction terms for each band. Instead, we estimate photometric zero points by directly bootstrapping against sources from the 2MASS PSC (Skrutskie et al. 2006). Although these are not standards, the ensemble of 2MASS stars may be treated as such. Since the disk of M31 is crowded, and 2MASS has low resolution ($1''$ pixel^{-1}), we choose to directly estimate zero points only in the sky images. We estimate the zero points of M31's images from a sliding window average of zero points from adjacent sky images (analogous to the median background subtraction procedure, described in Section 4.1.2).

Specifically, instrumental photometry of stars in the uncrowded sky fields is obtained with Source Extractor (Bertin & Arnouts 1996). We use the `AUTO` photometry mode to capture the full stellar light without using aperture corrections. 2MASS PSC objects are matched to our Source Extractor detections by position using J. Sick's *Mo'Astro*[10] Python package, which manages the full 2MASS PSC in a MongoDB database. The 2MASS PSC contains many galaxies, and many 2MASS sources are saturated in our deeper WIRCam images. Thus, we select sources with $J < 14$ or $K_s < 15$ magnitudes, and full width at half maximum $<1''$ according to our Source Extractor photometry. Additionally, we select sources with $J - K_s < 0.8$ (typical of foreground Milky Way stars) as we observe larger zero-point residuals in redder stars. After filtering, typically 200 matched 2MASS sources remain in typical WIRCam images. Given a joined catalog of 2MASS and instrumental photometry (in ADU) in a specific sky image, we estimate an instrumental zero point as the median photometric offset:

$$m_0 = \langle m_{\rm 2MASS} + 2.5\log_{10}({\rm ADU}/T_{\rm exp})\rangle. \quad (1)$$

Our data show no trend in zero-point versus $J - K_s$ color index. Hence, following practice at CFHT, we do not apply a color transformation between 2MASS and WIRCam bandpasses. Internal testing at CFHT with synthetic photometry indicates that color transformation coefficients may be $A_J = 0.05$ and $A_{K_s} = -0.005$ (K. Thanjuvar 2011, private communication). For typical M31 RGB stars with $J - K_s \sim 1$, this color transformation would be a <0.1 mag effect.

Given that 2MASS stars in each image have photometric uncertainties $0.05 \lesssim \sigma_{\rm 2MASSmag} \lesssim 0.3$, the typical statistical zero-point uncertainty, σ_{m_0}, is 0.1 mag in a single image. We reduce this random uncertainty to <0.01 mag by smoothing the zero-point time series with a sliding window average.

6. ANALYSIS OF SKY FLAT FIELDING AND BACKGROUND SUBTRACTION METHODS

Near-infrared sky flat fields are fraught with *additive* contaminants from thermal emission and scattered light. Although we regard a perfect near-infrared flat field as unattainable, we can test which flat-field prescription (`QRUN`, `NIGHT`, or `FW100K`) performs best and assess whether the final quality of our NIR M31 mosaics is limited by uncertainties from flat fielding or from ST-nodding background subtraction uncertainties.

6.1. Evolution of Real-time Sky Flats

A key advantage of sky flats is their close temporal correspondence to the data. Taken to the extreme, our `FW100K` flats are updated with sliding windows of approximately 30 minutes. Here we investigate the nature of evolution in the "real-time" `FW100K` flats throughout a night. In Figure 7 we show the evolution of `FW100K` sky flats relative to a single `NIGHT` sky flat over the course of 3 hr on a single night. This special sequence of engineering observations consisted of consecutive integrations on sky fields, without nodding to the M31 disk, so that an uninterrupted view of sky flat evolution could be visualized. Over the course of 3 hr, we see a large-scale shape perturbation move across the detectors from left to right. On the same timescale, the background *level* has changed by as much as 30% (Figure 7, middle panel).

Although Figure 7 clearly demonstrates that real-time `FW100K` sky flats evolve smoothly, it does not distinguish whether this evolution is driven by proportional effects or by additive contamination such as a thermal background or scattered light. However, we do note that the patterns are similar to those observed in the CFHT-IR camera by Vaduvescu & McCall (2004), who report a thermal background contamination.

An alternative interpretation is that these sky flat deviations are instabilities in the WIRCam detector electronics. The dominant macroscopic electronic feature in WIRCam flat fields are the amplifier bands. Each WIRCam detector is divided into 32 horizontal bands (each 64 pixels high) that are read out into independent amplifiers. These amplifiers have gains that result in levels that differ by 10% in flat-field images. However, we find that the gain of each amplifier band is stable throughout the night, at a level of $<0.1\%$ relative to other amplifiers. Thus, sky flat evolution is not driven by WIRCam gain instabilities, but by large-scale variations in the flat-field function or an additive background component.

6.2. Detector-to-detector Zero-point Evolution

We can test whether real-time sky flats are tracking evolution in the intrinsic WIRCam flat-field function or merely a background contamination, by examining the detector-to-detector photometric consistency against 2MASS standard photometry. Recall that our sky flats are designed to unify the zero points of the four WIRCam detectors by scaling according to the median background levels seen in each detector. Any additive background contamination will introduce detector-to-detector zero-point offsets. In Figure 8 we examine zero-point differences implied by the median background values in real-time `FW100K` sky flats, which are computed as $-2.5\log_{10}(\beta_i/\beta_1)$ from the discussion in Section 4.1.1. From Figure 8 we see that the estimated zero-point differences between detectors can be variable over a range of 0.05 mag in the K_s band. This variability is more prominent in K_s-band sky flats than in J-band ones.

We can test the validity of these zero-point transformations by verifying the photometric zero points of individual detectors against 2MASS stars, as was done in Section 5. Figure 9 shows the distribution of mean detector-to-detector zero-point offsets observed in images processed by `FW100K` sky flats. We find that zero points are consistent within ± 0.1 mag, though we detect a small possible systematic bias between detectors no. 1 and no. 4 at the level of 0.03 mag. The origin of this zero-point bias can be seen in the bottom panel of Figure 7, which tracks the detector-to-detector zero-point evolution estimated from each real-time `FW100K` sky flat over the course of 3 hr. The relative zero points slowly shift by $\lesssim 0.05$ mag in concert with the evolution in the shapes of `FW100K` sky flats due to time-varying additive contamination (upper panel of Figure 7). These results confirm that our sky flats *are* contaminated by a thermal background, albeit at a small level. The systematic photometric bias at the level of 0.03 mag is negligible compared to the photometric

[10] Publicly available at https://github.com/jonathansick/MoAstro.

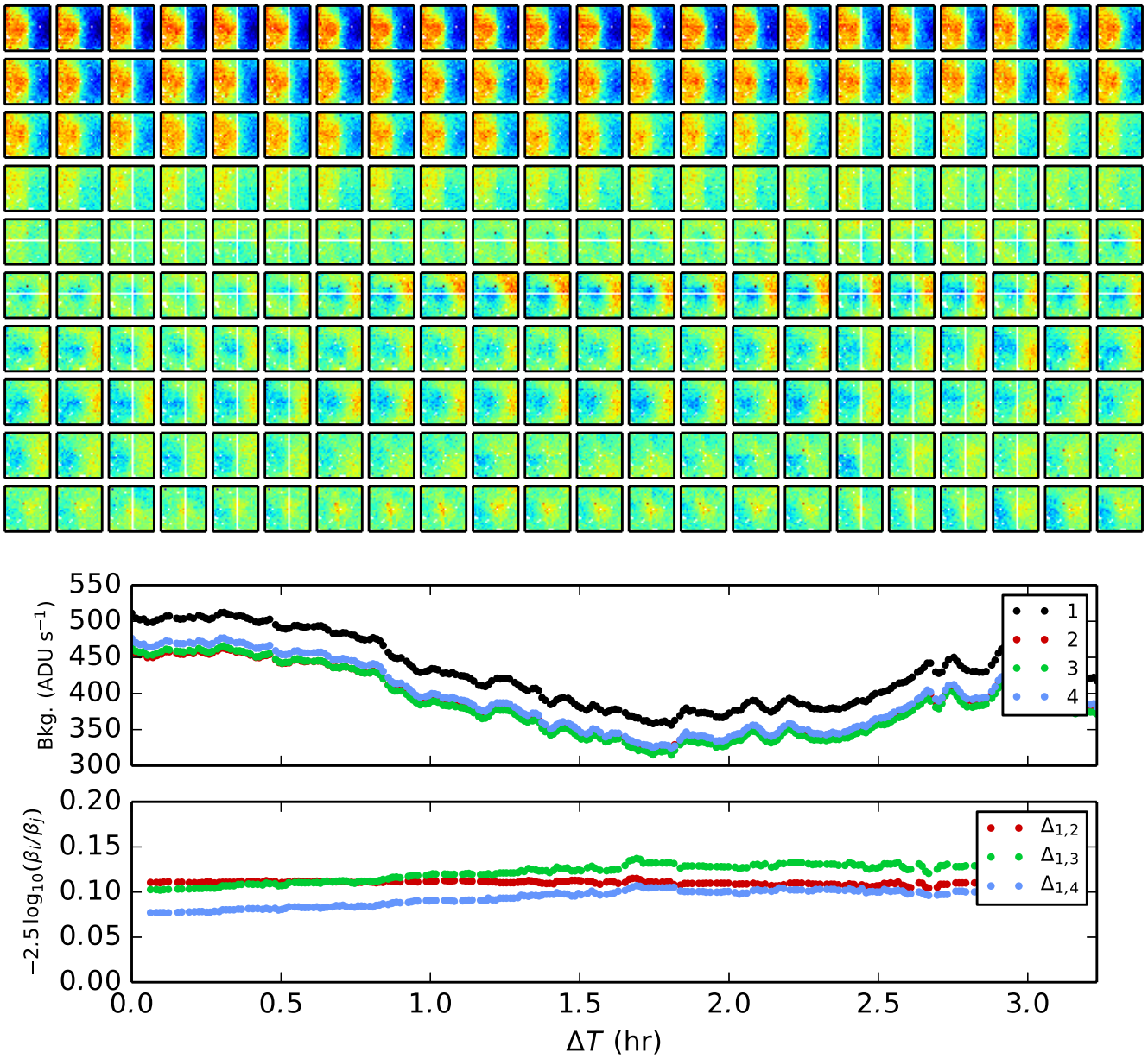

Figure 7. Evolution of FW100K K_s-band sky flats over the course of 3 hr. Percent difference maps of FW100K sky flats relative to the NIGHT sky flat are shown in the upper grid (time evolves left to right, from the top row). Colors in the percent difference maps show ±2% variation. The middle panel shows the background level observed in each detector as a function of time. The bottom panel shows zero-point differences computed for each FW100K sky flat as a function of time since the first FW100K flat between detector no. 1 and detectors nos. 2, 3, and 4, respectively.

(A color version of this figure is available in the online journal.)

uncertainty of individual stars. This result also suggests that our flat fields are significantly better than the ±4% systematic error upper limit established by comparing sky and dome flat fields (Figure 6).

6.3. Frame Residual Shapes

We have established (Sections 6.1 and 6.2) the presence of an additive contamination in WIRCam sky flats that varies over the course of a night and has a slight (<0.1 mag) influence on photometric calibration. Here we demonstrate how contamination in sky flats influences our observations of M31's surface brightness by examining the residual shapes of individual frames against the median shape of the disk (as assembled in our wide-field mosaic, Section 7). This also provides a test of the timescale over which the intrinsic WIRCam flat-field function is stable. If the residuals of data sets treated by QRUN or NIGHT sky flats vary systematically with time, in correspondence with the results of Section 6.1, then the flat-field function of WIRCam truly would be variable throughout a night. In this case, FW100K sky flats should be most appropriate. This effect should be exacerbated in signal-dominated (not sky-dominated) fields as flat-field errors grow in proportion to signal strength.

Our 2009B observations of the field M31–37 in the K_s band are ideal for this experiment: a single detector in that field covers the core of M31, and observations were taken in two blocks, covering a total window of 2 hr (most blocks for this program are observed by the CFHT queue in a half hour). Both the high surface brightness and wide time baseline of this field should highlight flat-field bias and variation. In Figure 10 we show the residual shapes of individual WIRCam frames against the median shape of the mosaic, given FW100K, NIGHT, and

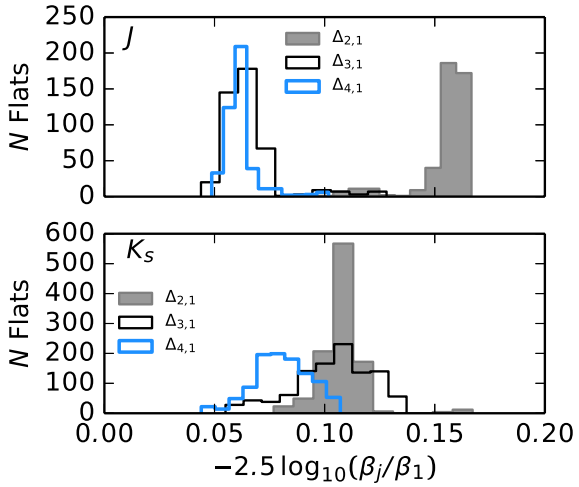

Figure 8. Distribution of real-time sky flat scaling factors, measuring detector-to-detector zero-point differences relative to detector no. 1 (gray: no. 2, black outline: no. 3, blue outline: no. 4) in J and K_s bands.

(A color version of this figure is available in the online journal.)

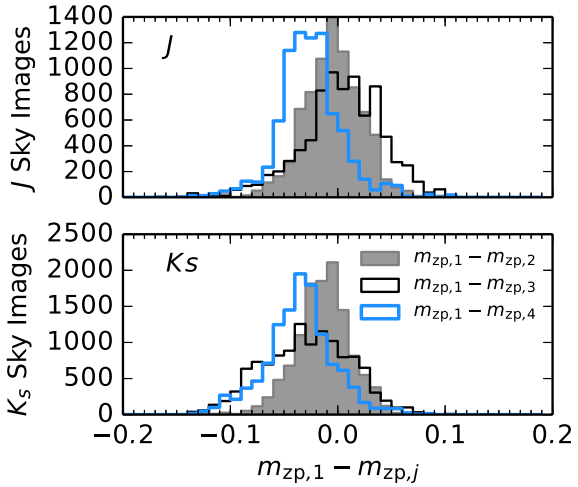

Figure 9. Distribution of mean detector-to-detector zero-point offsets for sky images processed by real-time sky flats. Zero-point offsets between detectors no. 1 and no. 2, no. 1 and no. 3, and no. 1 and no. 4 are plotted as gray, black outlined, and blue outlined histograms, respectively, for the J band (top) and K_s band (bottom).

(A color version of this figure is available in the online journal.)

QRUN sky flattening of K_s-band images at the M31–37 field. To analyze the shapes of these difference images, we marginalize along their rows (left side of Figure 10) and columns (right side of Figure 10). Note that these marginalizations are done for each detector in the 2 × 2 WIRCam array; the core of M31 resides in detector no. 2 (lower right). In that high surface brightness region, there are strong surface brightness residuals that clearly point out flaws in the flat field itself.

Comparing panels in Figure 10, we see that both FW100K and NIGHT sky flats have similar performance, where frames vary in surface brightness by ±0.5% at the core of M31. The exception are QRUN-treated frames that show an evolution on the order of ±1% of the K_s-band sky brightness over a similar timescale as indicated in Figure 7. This indicates that QRUN sky flats, which are built over several nights of data, are unsuitable for capturing the WIRCam flat-field function. While this indicates that the intrinsic WIRCam flat-field function varies detectably from night to night, the performance equivalence of FW100K and NIGHT sky flats indicates that the WIRCam flat *does not* vary throughout the night.

It is useful to contrast the frame shape residuals seen in detector no. 2 with those in other detectors, where the disk surface brightness is lower. There, FW100K (Figure 10(a)), NIGHT (Figure 10(b)), and QRUN (Figure 10(c)) show similar residual distributions, on the order of $\lesssim$0.2% of the NIR background brightness. Further, the results are not monotonically varying in time, as they are in detector no. 2, and indeed appear to vary essentially randomly. We interpret this behavior as being caused by random additive background processes, distinct from flat-field biases that are proportional to surface brightness. We extend this analysis across the entire data set below.

6.4. Distributions of Frame Shape Residuals

Figure 11 shows the distribution of frame-block residual shape amplitudes, measured at the 95% difference interval to quantify the reliability of recovering surface brightness shapes in individual WIRCam frames. As in our test of median sky frame flatness (Section 6.5), we see that the consistency of frame shapes is ∼0.3% of the background level. This result is seen universally among the QRUN, NIGHT, and FW100K sky flat pipelines and for 2007B and 2009B observing schemes, agreeing with our observation in Section 6.3 that in background-dominated regimes frame shape consistency is *not* correlated with flat-field bias. Rather, we interpret Figure 11 as measuring the amplitudes of *additive* stochastic background shapes either originating from the sky or associated with the instrumentation itself. Effectively, Figure 11 illustrates the *flatness limit* of WIRCam frames observed with large ST nods, sky flat fielding, and median sky subtraction.

6.5. Shapes of Median Background Frames

Another test of sky flats is their ability to produce an unbiased sky background, up to the level of intrinsic sky variations. This test can be made by examining amplitude of median background frames (Section 4.1.2) produced by QRUN-, NIGHT-, and FW100K-processed data sets. We measure the amplitude of shapes across the 10′ × 10′ WIRCam frame as the 2-standard-deviation interval (95%) of each median sky image's pixel distribution: 2σ(medsky). In Figure 12 the cumulative distribution functions of the background shape amplitudes are presented for each set of flat-fielded data, in each band and for each semester.

QRUN sky flats, which are known to be incorrect (Section 6.3), produce median backgrounds with the largest amplitudes—as much as 3% of the background. Compared to the other flats, the QRUN sky flats are biasing the background shape, relying upon the median background subtraction process to effectively flatten the sky background. FW100K sky flats have the opposite effect. The temporal windows from which FW100K sky flats are constructed have nearly the same span as those for median background images, so the background amplitude is unsurprisingly (and unrealistically) flat. Thus, FW100K sky flats are incorrect as they divide all structure in images, be they multiplicative or additive in origin. NIGHT flats produce median backgrounds with moderate, realistic amplitudes.

A surprising result from Figure 12 is that backgrounds in 2009B images are not flatter than in 2007B. Recall that median sky images are composed of five sky frames taken closest to a disk frame. In 2007B, all sky frames were sampled from the same coordinate on the sky and span a 12 minute window covering sky integrations taken before and after a disk image (for both J and K_s ST nods). In 2009B, sky frames were sampled from randomly chosen sites along the sky field ring (Figure 1) with a

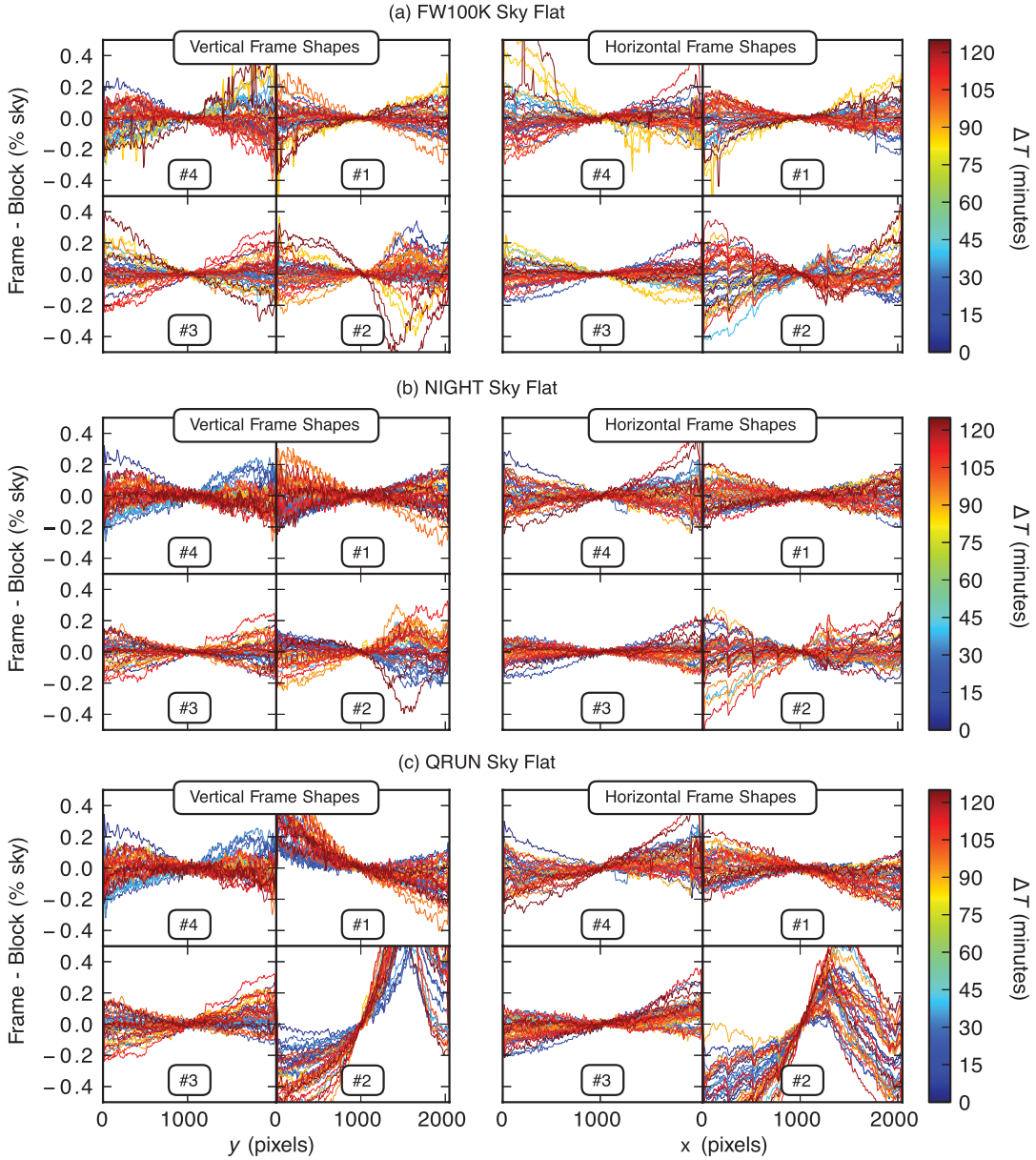

Figure 10. Comparison of frame shape residual variations for images processed with (a) `FW100K`, (b) `NIGHT`, and (c) `QRUN` sky flats. Residuals are the difference of an individual frame to the median (mosaic) shape. Images are from the M31–37 field, K_s band, observed in the 2009B semester. Residuals have been marginalized across the x (left) and y (right) axes to provide one-dimensional views. Axes match WIRCam's 2×2 detector footprint. Individual integrations are colored by their time after the first disk integration. The center of M31 is located in the lower-right detector (no. 2); surface brightness bias in these regions betrays the presence of flat-field bias. Lower surface brightness regions are dominated by shape variations on the order of $\pm 0.2\%$ of background level, interpreted as additive uncertainties associated either with the detector, skyglow, or both. `NIGHT` sky flats reliably capture the disk shape in the signal-dominated detector no. 2 as well as, if not better than, `FW100K` flats. Hence, the WIRCam flat-field function is stable over a night. `QRUN` sky flats introduce large biases in the bulge-dominated surface brightness in detector no. 2 (lower right). In the more sky-dominated regions of the image (detector no. 4), `QRUN` sky flats produce images with similar stability to `FW100K` and `NIGHT` sky flats, indicating that the limit of additive uncertainties associated with sky or instrumental background variations is reached here.

(A color version of this figure is available in the online journal.)

window typically spanning 15 minutes. Thus, both 2007B and 2009B median sky images span similar time windows, although the 2009B strategy attempts to marginalize over five distinct sites on the sky (and thus sky background shapes) while 2007B median sky images do not. From Figure 12 we conclude that this marginalization does not effectively occur, implying that the background shapes sampled from distinct sites on the sky *are* correlated. Since wide-field movies of the NIR sky (Adams & Skrutskie 1996) suggest that this should not be true, much of the structure in the median background images is an instrumental background. Similar backgrounds are seen by Vaduvescu & McCall (2004) in the CFHT-IR camera.

6.6. Section Summary

In this section we have analyzed the performance of our flat-field, median background subtraction and photometric calibration procedures outlined in Figure 4. Here we summarize our findings on the accuracy of surface brightness shapes reproduced by WIRCam in an ST nodding observing program.

Night sky flat fields are preferable to dome flats because sky flats properly calibrate dust, detector surface, and gain structures thanks to proper illumination of the WIRCam entrance pupil. Such flats can be easily made from program data. Specifically, we advocate `NIGHT` flats, which are built from all sky images

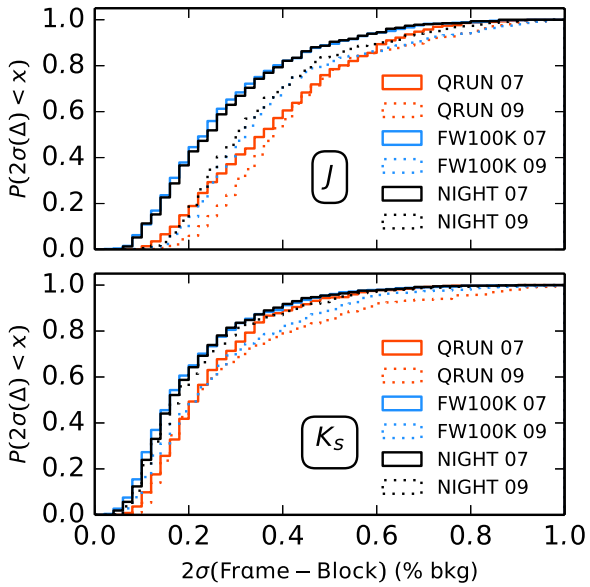

Figure 11. Cumulative distributions of scalar difference amplitudes between individual frames and blocks in the J (top) and K_s (bottom) mosaics, measured as a dispersion of pixel differences at the 95% level. Whether processed with QRUN (orange), NIGHT (black), or FW100K (blue) sky flats, or observed in 2007B (solid lines) or 2009B (dotted lines), the residual amplitude differences between frames and blocks are similarly distributed. The mean amplitude difference is 0.3% of the J background brightness (0.2% in K_s).

(A color version of this figure is available in the online journal.)

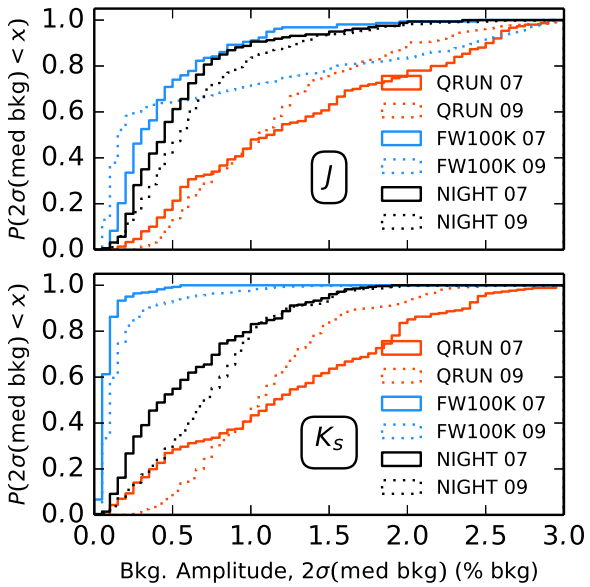

Figure 12. Cumulative distribution function of background level amplitudes across median sky images processed with QRUN (orange), NIGHT (black), and FW100K (blue) sky flats for the J (top) and K_s (bottom) bands. Real-time (FW100K) sky flats produce much flatter median sky frames, with mean shape amplitudes of 0.3% (J) to 0.1% (K_s) of the NIR sky level, while the mean amplitude of QRUN-flat processed sky images is 1.5% of the background level, and as high as 3% of the background level. NIGHT flat-processed data generate mean shape amplitudes of ∼0.6% of the background level. We interpret this as QRUN flats introducing a bias in background shapes, while FW100K flats create unrealistically flat backgrounds by dividing skyglow structure that should be left for background subtraction. NIGHT flats have realistic background amplitudes that are similar in both bandpasses.

(A color version of this figure is available in the online journal.)

taken over a single night, because they capture night-to-night variations in the WIRCam flat field not captured by QRUN flats (compare Figures 10(b) and (c)).

Although we confirm that night-sky flats are afflicted by an *additive* contamination (i.e., thermal background, or scattered light) that varies on sub-hour timescales (Figure 7), this contamination impacts the photometric zero point at a level of 0.03 mag. This *does not* impact the overall surface brightness fidelity of our mosaic. Since the ANDROIDS/WIRCam mosaic of M31 is almost entirely background dominated, we are principally limited by background subtraction, which is non-trivial since the background cannot be directly measured on the M31 disk given our current field of view. Ultimately we find that the *shape* of the background on the disk can be known to within 0.3% of the NIR background levels (Figure 11). By comparison, the typical amplitude of median sky images is ∼0.6% of both the J and K_s background levels. Note that it is impossible for an NIR observing program with large ST nods to subtract background shapes better than the 0.3% we find here: the shape of the background at the target will always be distinct from background shapes measured at designated sky fields.

To summarize the first half of this work, we have demonstrated that NIGHT sky flats are appropriate for our application, and the impact of thermal background and scattered light on flat fields and zero points is negligible compared to background uncertainties due to NIR skyglow. In the following sections we assemble NIR mosaics of M31 using images processed according to the preceding sections, using NIGHT sky flats, and show that the single most important calibration for wide-field NIR mosaicking is sky offset optimization.

7. SKY OFFSET OPTIMIZATION

In Figure 13(a), we plot mosaics (assembled using *Swarp*) from ANDROIDS frames processed with the pipeline discussed in Sections 3–5 and adopting the NIGHT sky flat prescription favored in Section 6.6. Although these image preparations can constrain the surface brightness *shape* of a WIRCam frame by $\lesssim$0.3% of the sky level, Figure 13(a) demonstrates that the true level of the background on M31's disk is lost by the temporal and spatial variation of skyglow between disk and sky field observations. This results in the distinct field-to-field surface brightness discontinuities seen in Figure 13(a). However, we can use the constraint that all overlapping pairs of images composed in our mosaic should have equal surface brightness in their intersections. To enforce this constraint, we solve for a *sky offset* for each image: a small scalar nudge of intensity that can be added or subtracted from each image so that all images in the mosaic have continuous surface brightnesses. Note that we use the term sky offset for these intensity adjustments, but in practice these offsets are agnostic of the cause of background subtraction error that they correct. Since our mosaics are made from many inter-connected images (3924 J and 4972 K_s image frames), our optimization of sky offsets can provide powerful constraints on the true level of the background at the M31 disk.

Montage is a FITS mosaicking package (Berriman et al. 2008) originally written for 2MASS that includes sky offset estimation (background rectification, in their terminology) functionality. *Montage* can solve sky offsets either as scalar levels or as planes. Sky offsets are then chosen iteratively by looping through each image pair and choosing the offset needed to minimize the difference image of that pair, counting previous sky offset estimates. Sky offsets are refined over several loops through the entire set of overlapping image pairs until convergence is reached (that is, once incremental adjustments to sky offsets diminish below user-specified threshold). Although this iterative implementation of sky offset optimization is elegant, its accuracy has never

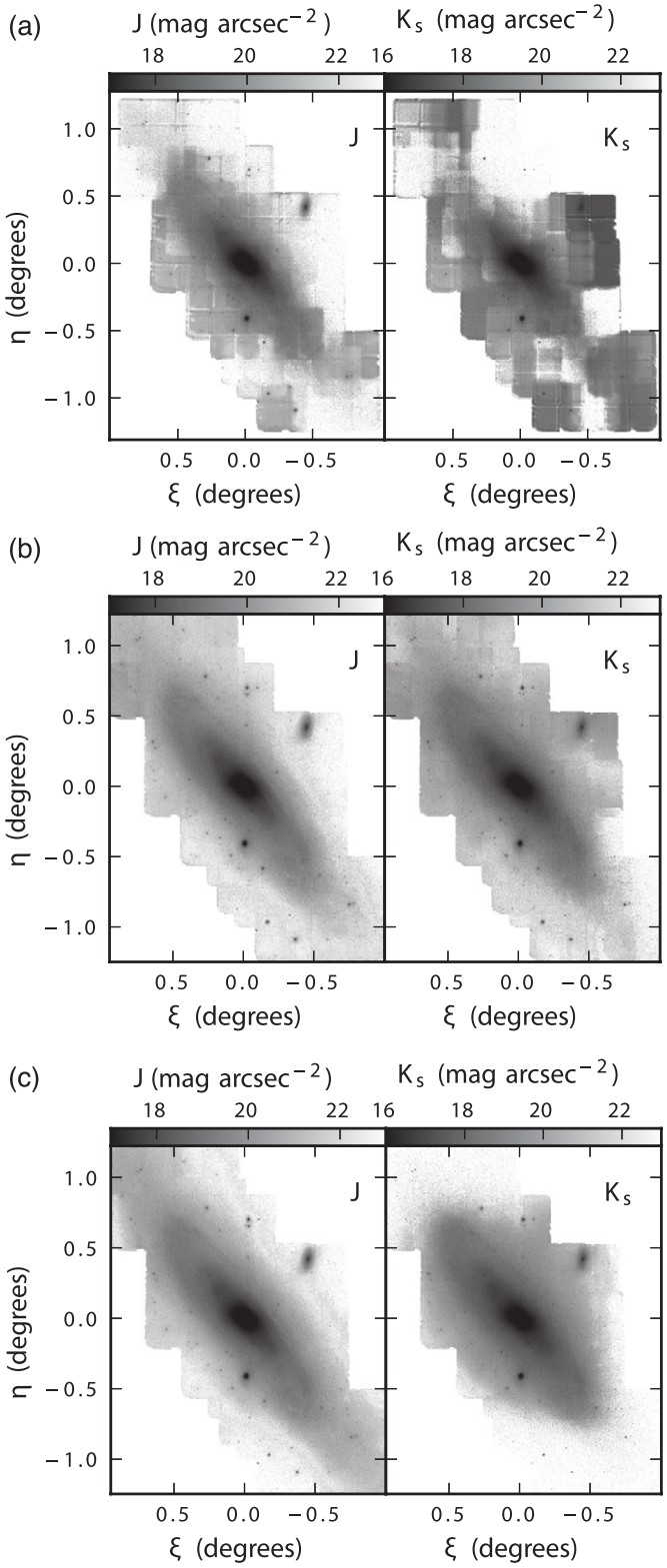

Figure 13. ANDROIDS/WIRCam mosaics of M31 in J (left panels) and K_s (right panels). (a) Median background-subtracted mosaics processed according to Sections 3–5 using NIGHT sky flats, (b) mosaics after scalar sky offsets are applied, as described in Section 7, and (c) mosaics generated by *Montage* using planar sky offsets.

been formally analyzed in literature, to our knowledge. In particular, we are interested in the robustness of Montage sky offsets against local minima in the N-dimensional solution space of sky offsets, given a mosaic of N independent images. Further, the optimization is slow, given the several thousand frames in our mosaics. Thus, we decided to implement our own sky offset algorithm,[11] although a comparison to the *Montage* solution is given in Section 9.

7.1. Sky Offset Implementation

Our sky offset algorithm is based on two features that distinguish it from the *Montage* implementation. First, the optimization is carried out in three hierarchical stages to accommodate the large number of images. Second, we use a downhill simplex algorithm (Nelder & Mead 1965, hereafter NM) with re-convergence checks rather than the iterative approach of the *Montage* sky offset solver. We begin with NIGHT-sky flat calibrated, median sky subtracted, and photometrically calibrated image sets that are resampled using *Swarp* to a common pixel in an Aitoff equal-area project with the native WIRCam pixel scale of $0\farcs3$ pixel^{-1}.

We address the sky offset optimization hierarchically by considering the geometry of the WIRCam detectors in a 2×2 grid and arrangement of 39 WIRCam fields on the M31 disk (see Figure 1). The first stage of optimization is to stack all *detector frames* (images taken with a given detector, at a given field) into *stacks*. The offsets applied to WIRCam frames to build stacks are labeled Δ_F. Next, we solve for the offsets Δ_S to ensure surface brightness continuity across the four stacks in a WIRCam field. We call the combined unit of four stacks a *block*. The last stage of optimization solves for the offsets Δ_B applied to each block to ensure surface brightness continuity across the mosaic. The net scalar sky offset applied to each frame is thus

$$\Delta_\Sigma = \Delta_F + \Delta_S + \Delta_B. \quad (2)$$

After each stage of optimization, the scalar sky offsets are added to the WIRCam frames, stacks, and blocks and *Swarp* is used to median co-add the images to generate stacks, blocks, and a mosaic, respectively. Using this hierarchical scheme ensures that, at worst, the number of dimensions in our optimizations is 39 as opposed to the number of WIRCam frames (a factor 10^2 reduction).

We use two algorithms for solving sky offsets. Solving Δ_F offsets in the first stage is trivial since all frames simultaneously overlap. Thus, it is sufficient to simply compute a mean surface brightness across all frames and directly compute offsets (Δ_F) between the levels of each frame and the mean level. In the last two stages, stacks and blocks, respectively, are arranged in networks of overlapping pairs. For this case we introduce our NM simplex-based offset optimization algorithm.

We identify overlaps between images in a brute-force fashion according to their frames in the mosaic pixel space, defined by the CRPIX, NAXIS1, and NAXIS2 header values of the resampled images. For each overlapping image pair, we compute a difference image, and ultimately a median difference, $\langle \boldsymbol{I}_i - \boldsymbol{I}_j \rangle$. While computing the median difference, we mask bad pixels using weight maps (propagated by *Swarp*) and expand this mask with sigma clipping. Along with a difference estimate, we also record the area A_{ij} of unmasked pixels in the overlap and the standard deviation of the difference, σ_{ij}.

We can estimate the optimal set of scalar sky offsets, Δ_i, for each image i by minimizing the objective function:

$$\mathcal{F}(\Delta_1, \ldots, \Delta_n) = \sum_{i,j} \mathcal{W}_{ij}(\langle \boldsymbol{I}_i - \boldsymbol{I}_j \rangle - \Delta_i + \Delta_j)^2. \quad (3)$$

[11] Publicly available at https://github.com/jonathansick/skyoffset.

Note that each coupled image pair is its own term in the objective summation, and that there are as many degrees of freedom (Δ_i) as there are images in the mosaic. Each coupling is tempered by a weighting term $\mathcal{W}_{ij}$:

$$\mathcal{W}_{ij} = \frac{A_{ij}}{\sigma_{ij}}, \quad (4)$$

so that more priority is given to couplings of larger areas (A_{ij}) and small standard deviations of their difference images (σ_{ij}).

The objective function in Equation (3) puts no constraint on the net sky offset: $\sum \Delta_i$. Assuming that background subtraction errors are normally distributed, and not biased, sky subtraction offsets should not add a net amount of flux to the mosaic. Fortunately, it is possible to impose this constraint post facto by subtracting the mean offset from the sky offsets:

$$\Delta_i^* = \Delta_i - n^{-1} \sum_{j=1}^{n} \Delta_j. \quad (5)$$

In the limit that sky offsets Δ_i are drawn from a Gaussian distribution, with standard deviation σ_Δ, the absolute brightness of the whole mosaic will be uncertain by $\sigma_\Delta/\sqrt{N_{\rm images}}$. The consequences of this uncertainty are revisited in Section 9.

Given the image coupling records, we optimize the set of Δ_i by applying the object function (Equation (3)) to the NM downhill simplex algorithm. The NM algorithm is naturally multidimensional and does not require knowledge of the gradient of the objective function. Instead, the NM algorithm operates by constructing a geometric simplex of $N+1$ dimensions that samples the sky offset parameter space. By evaluating the objective function at each vertex of the simplex, the NM algorithm adapts the simplex shape to ultimately contract upon a minimum.

The NM algorithm will converge into any local minimum without necessarily seeking the global minimum of the objective function. We resolve this issue with two methods: ensuring reconvergence, and sampling different starting conditions.

The practice of ensuring reconvergence in a downhill optimization is suggested by Press et al. (2007). Upon each convergence, the optimal point in the simplex, $\boldsymbol{p}$, is recorded. A new simplex is then generated where one vertex is $\boldsymbol{p}$ and the rest are $\boldsymbol{p} + \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is a normal random variable of mean zero and standard deviation $\sigma_{\rm restart}$. That is, the simplex of the restart retains one vertex upon the previously found minimum, while the other vertices surround that minimum. We set $\sigma_{\rm restart}$ to 2× the dispersion of image-to-image differences. Our optimization iterative converges and re-converges simplexes until the same minimum is consecutively arrived upon, indicating that the NM algorithm has arrived upon a robust solution. Our sky offset optimizations for 39 blocks typically require ∼1000 restarts before converging definitively.

Besides ensuring reconvergence, we also start several independent NM simplex optimizations from random starting points in parameter space to seek a globally optimal sky offset solution. We find that $N_s = 50$, and possibly fewer, starts are quite sufficient for an optimization with 39 sky offset parameters (such as the fitting of Δ_B block offsets in mosaic). For each start, an initial simplex is generated randomly. Since each point in the N by $N+1$ simplex is a suggested sky offset for a given field, each offset is randomly sampled from a normal distribution whose dispersion is 3× the standard deviation of image-to-image differences to ensure that parameter space is well covered. Note that each simplex start and series of subsequent restarts can be performed in parallel. Once all simplex runs are complete, the set of sky offsets belonging to the run that yielded the smallest value of the objective function is adopted.

Table 2
Hierarchy of Scalar Sky Offsets (Using NIGHT Flat Fielding, and Median Background Subtraction)

Offset Type	Sem.	J $\frac{\sigma_\Delta}{\langle I_{\rm bkg}\rangle}$ (%)	K_s $\frac{\sigma_\Delta}{\langle I_{\rm bkg}\rangle}$ (%)
Δ_F	07B	2.54	2.29
	09B	1.88	1.87
Δ_S	07B	0.08	0.05
	09B	0.05	0.03
Δ_B	07B	1.25	0.94
	09B	0.70	1.14
Δ_Σ	07B	2.73	2.44
	09B	1.98	1.88

Notes. Each level of sky offset is defined in Equation (2). $\langle I_{\rm bkg}\rangle$ is taken as the instantaneous background level for the images being sampled.

8. ANALYSIS OF SCALAR SKY OFFSETS

Figure 13(b) presents the fruits of our WIRCam pipeline and sky offset optimization. Compared to our mosaics without sky offsets, Figure 13(a), the sky offset optimization is clearly essential for assembling wide-field NIR mosaics. These mosaics are not yet perfect; field-to-field discontinuities at a level of 0.05% of background remain, and large-scale background residuals perturb the outer M31 disk.

8.1. Amplitudes of Sky Offsets

The distribution of scalar sky offsets provides an excellent characterization of background subtraction uncertainties when using ST nodding. Recall that sky offsets are optimized hierarchically: WIRCam frames are fitted to stacks, stacks are fitted into blocks of four contemporaneously observed WIRCam detector fields, and these blocks are fitted into a mosaic. Table 2 lists the standard deviations of these offset distributions with respect to the typical background level observed in the J and K_s bands.

Note that the sky offsets, as a percentage of background level, are comparable in the J and K_s bands, despite the background being ∼4× brighter in K_s than J. This indicates that spatio-temporal variations in the NIR background are monochromatic.

Within the hierarchy of background fitting, simply fitting frames to a stack (with Δ_F) is a correction on the order of 2% of the background intensity. Fitting blocks into a mosaic (Δ_B) is a further ∼1% correction. Overall, the temporal and spatial lags of ST nodding induce a 2% uncertainty in the background level at the target (see Table 2). It is this level of uncertainty that sky offset optimization must diminish to transform uncorrected mosaics (Figure 13(a)) into ones that reproduce the disk with fidelity (Figure 13(b)).

Note that offsets to fit a stack into a block (Δ_S) of four detector field stacks are smallest: 0.1% of the background level. This suggests that on the scale of the 2 × 2 WIRCam array, the contemporaneously observed detector frames are subjected to nearly identical biases in background. Stack offsets, then, arise from uncertainties in the pipeline's measurement of the background level from single frames in two stages: estimating

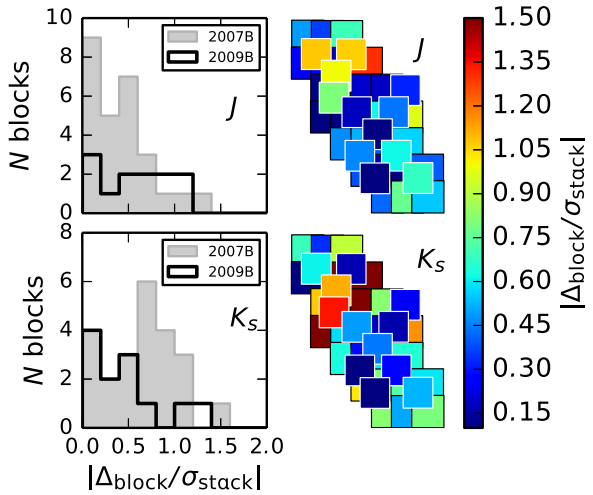

Figure 14. Acceptability of J and K_s scalar sky offsets between blocks, as measured by the ratio of $\Delta_B/\sigma_{\Delta_F}$, plotted as histograms and field maps. Shaded and black-outlined histograms distinguish blocks observed in 2007B and 2009B, respectively. Scalar sky offsets required for blocks are consistent with the background level uncertainties of single frames, given sky-target nodding background subtraction.

(A color version of this figure is available in the online journal.)

detector-to-detector zero-point offsets from frame background levels (Section 4.1.1) and again when subtracting a median background frame (Section 4.1.2). Indeed, in Section 6 we showed that median background images have shape amplitudes of 0.3% of the background level and that individual frames have surface brightness shapes that are uncertain at a level of 0.2%; Δ_S sky offsets are thus a consequence of the limited surface brightness flatness across a WIRCam frame.

8.2. Acceptability of Sky Offsets

Recall that scalar sky offsets were initially introduced as intensity increments to overcome uncertainty in the background level of detector field stacks. For sky offsets to be considered acceptable, we demand that the offsets applied to blocks, Δ_B, be consistent with the background level uncertainty of the blocks themselves. We can conservatively measure the background uncertainty as the dispersion of Δ_F frame offsets in a stack: σ_{Δ_F}. If sky offsets fitted between blocks are statistically permissible, then $\Delta_B \lesssim \sigma_{\Delta_F}$. In Figure 14, we plot field maps (in the same spatial configuration as Figure 1) painted with the values of $\Delta_B/\sigma_{\Delta_F}$ for each block in the J and K_s mosaics. The sky offsets are indeed distributed within the uncertainty budgeted by σ_{Δ_F}: the sky offsets are statistically acceptable.

One can also see the veracity of these sky offsets by plotting a time series of both directly measured background levels and background levels interpolated on disk observations via sky offsets. In such plots we see remarkable continuities of the background level as estimated from sky offsets. Through the ST nodding and sky offset optimization, we *have* effectively measured the background level on M31.

8.3. Residual Image Level Differences

Although scalar sky offsets are statistically valid, they are imperfect prescriptions against the background subtraction uncertainties of each image stack—that much is visually true. We quantify the limited effectiveness of sky offset fitting as the image differences between coupled blocks i and j, $(\boldsymbol{I}_i - \Delta_{B,i}) - (\boldsymbol{I}_j - \Delta_{B,j})$, after the sky offsets $\Delta_{B,i}$ have been optimally fitted to each block. Table 3 lists distributions of both image level differences before and after the application of scalar sky offsets. Uncorrected, the ensemble of coupled blocks have a mean intensity difference of ∼1% of the typical background intensity. Scalar sky offsets decrease the differences between overlapping fields to ∼0.2%.

Table 3
Coupled Block Intensity Differences and Residual Intensity Differences After Application of Scalar Sky Offsets: 25th, 50th, and 75th Percentiles of Distribution

	Coupled Block $\langle I_i - I_j\rangle/\langle I_{\rm bkg}\rangle$ (%)		
	25th	50th	75th
J, uncorrected	0.49	0.91	1.76
J, scalar offset	0.05	0.10	0.17
K_s, uncorrected	0.44	0.91	1.41
K_s, scalar offset	0.02	0.04	0.08

Note. Differences are presented as a percent of the mean background level seen by observations in each band.

Figure 15(a) shows the block-to-block residual differences as a fraction of the local surface brightness. Note that throughout the bright inner disk of M31, block-to-block residuals are negligible compared to the disk signal; at the mosaic periphery ($R \sim 20$ kpc), field-to-field residuals become comparable to, or greater than, the disk surface brightness. The poor fit is driven primarily by diminishing disk signal, rather than poor convergence of sky offsets. This can be seen by plotting the magnitude of block-to-block residuals (in units of background brightness) in Figure 15(b). There, significant residuals are distributed throughout the disk, rather than the low surface brightness periphery of the mosaic.

The inability of scalar sky offset optimization to eliminate residual image differences should not be interpreted as a failure to detect the global minimum; the sky offset optimization algorithm (Section 7.1) appears robust in yielding this offset solution set. Evidence of this can be seen in Figure 15(c), where block-to-block network connections are colored by the ratio of the residual block-to-block intensity difference to the uncertainty in the block-to-block difference image. The sky offsets solved by the NM simplex algorithm are within the uncertainties of the difference images themselves; better scalar sky offsets *cannot* be made with the WIRCam blocks that our pipeline has produced. Our ability to produce a continuous NIR mosaic is fundamentally limited by our ability to subtract the true background shape seen at the M31 disk. As described in Section 6, ST nodding on the scale of M31 introduces an intrinsic shape uncertainty of 0.3% of the NIR background levels.

8.4. The Growth of Sky Offsets in Time and Space and Effectiveness of the 2009B Strategy

In Sections 8.1–8.3, we established the usefulness of sky offsets. We now address the value of the 2007B and 2009B observing programs, with the ST nodding observing strategies (introduced in Section 2). A key question is whether the 2009B strategy to minimize background level uncertainty by minimizing the latency between sky and target samples to ∼1.2 minutes is justified. Overall the 2009B semester reduced sky offsets by 30% (Table 2), at the cost of a roughly 50% reduction in observational efficiency (see Table 1). Here we show that the background certainty of any ST nodding campaign is limited both by this minimum sky sampling latency and also by the spatial structure of the sky background.

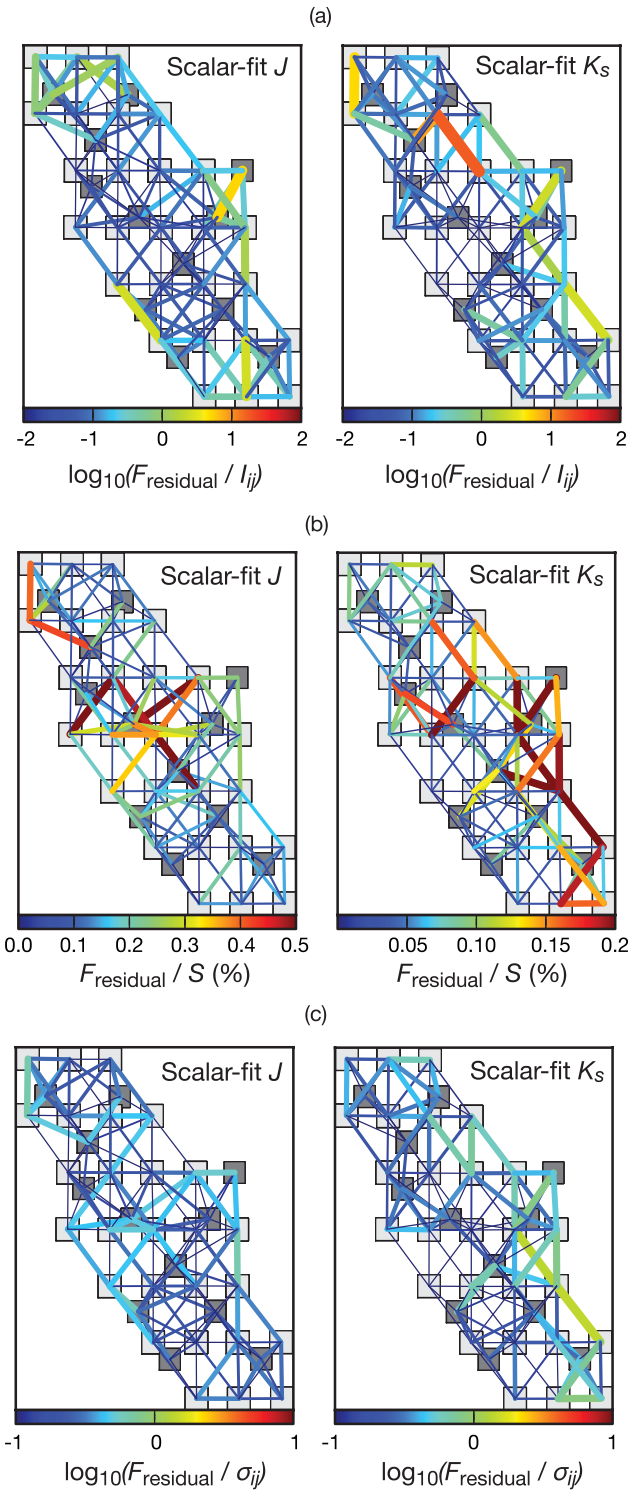

Figure 15. Map of residual block-to-block surface brightness differences after sky offsets: (a) as a fraction of the mean local surface brightness, (b) as a fraction of background level, (c) as a fraction of the standard deviation of the difference image. Thicker lines denote larger residual differences between overlapping fields (see also the color mapping). These graphs mimic the spatial distribution of the 2007B and 2009B WIRCam fields (Figure 1), with the footprints exploded to allow room for lines to connect coupled blocks. Fields observed in 2009B are plotted as darker squares than those observed in 2007B.

(A color version of this figure is available in the online journal.)

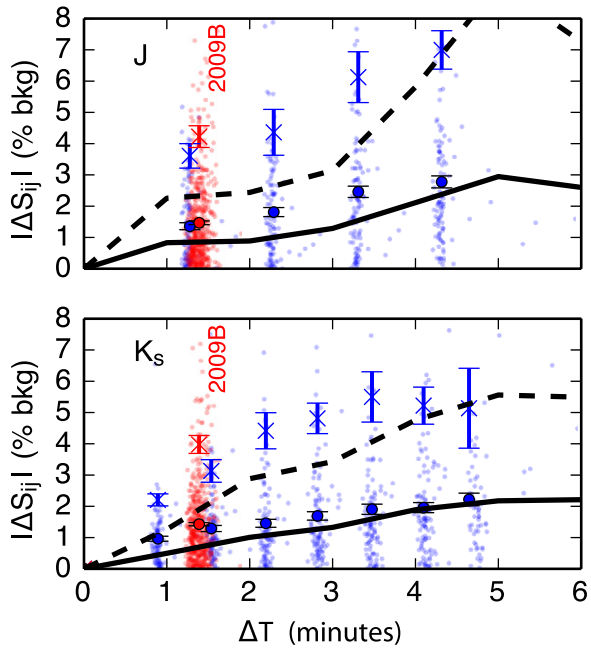

Figure 16. Temporal growth of background level variations observed at stationary sites on the sky vs. with sky-target nodding. Solid and dashed back lines mark the mean and 95% levels, respectively, of background variation observed in a single field without telescope nodding. Small blue and red dots show the net sky offset levels in 2007B and 2009B, effectively indicating the variation in background sky level seen as the telescope is nodded. Mean and 95% levels of sky offsets in the 2007B semester are plotted as large blue circles and crosses, respectively, while the same for the 2009B semester is plotted as red symbols, where sky latency was constrained to 1.2 minutes in both bands.

(A color version of this figure is available in the online journal.)

To test and distinguish temporal and spatial variations in the sky background, we plot the growth of background level variations versus time in Figure 16. As a fiducial for the intrinsic behavior of background variations, we measure the mean and 95% background change as a function of time at a stationary site on the sky (that is, a single sky field, without telescope nodding). In agreement with Vaduvescu & McCall (2004), we see a mean background level variation of ∼0.5% in 1 minute for both J and K_s bands. After 5 minutes, the intrinsic background level variation typically grows to 2%. At worst, we see background variations (measured at the 95% level of the sample distribution) of 5% in 5 minutes.

Individual points in Figure 16 are net sky offsets of disk images plotted against the time latency to the paired sky sample. The periodic time structure in Figure 16 is a consequence of the 2007B and 2009B ST nodding schemes (see Table 1); the circles and crosses in Figure 16 denote the mean and 95% level of background variation, respectively, in each cluster. Recall that all 2009B disk integrations have equal sky sample latency due to the *ST–target-sky* nodding pattern. In both J and K_s bands, we see that nodding the telescope between sky and target *generates* additional background level uncertainty beyond that expected from strictly temporal sky background evolution. This makes sense in the context of spatial sky variations (Adams & Skrutskie 1996). As shown in Figure 16, the process of ST nodding can inflate sky background variations by 1.5–2 times the background variability expected at a stationary site on the sky on short timescales. On longer timescales, the nodding and stationary sky variance converge, perhaps indicative of the timescales that NIR skyglow structures move across a nodding distance (1°–2° on the sky).

This analysis underscores the challenge of accurately recovering surface brightness in a wide-field NIR mosaic. ST nodding with CFHT implies typical time latencies of 60–70 s and

nodding distances of 1°–2°. Both of these elements prevent the true level of the background on M31's disk, in any single frame, from being known to an accuracy greater than 2%.

This also highlights why the 2009B program could only reduce the distribution of sky offsets by 30%. The rather shallow slopes of the mean background variance seen in ST nodding demonstrate the modest gain in background certainty by capping sky sampling latency at 1.2 minutes in 2009B compared to allowing latencies of 5 minutes in 2007B (see also Table 2). We also note that the expected 2009B sky offsets, at ∼1.2 minute latency, are 10% larger than those from 2007B in both J and K_s bands. While this could indicate different physical behaviors in the background between the 2007B and 2009B semesters, we also note that the nods employed in 2009B were larger than in 2007B (Figure 3). Had the 2009B campaign used the same sky fields as in 2007B, rather than the pseudo-random sky ring, the performance of the 2009B sky offsets listed in Table 2 could be more impressive. Ultimately, future NIR observers with similar large ST nodding programs may choose to implement 2007B- or 2009B-like observing strategies depending on the priority of observational efficiency or absolute certainty of the background level. Figure 16 should be useful in planning such programs.

9. SYSTEMATIC UNCERTAINTIES IN SURFACE BRIGHTNESS RECONSTRUCTION

Sky offsets produce a mosaic that is rigorously optimal only in the sense of field-to-field surface brightness continuity—not absolute background subtraction. In this section, we attempt to gauge the systematic surface brightness error inherent in the sky offset technique.

9.1. Comparison to Montage-fitted Images

Besides the simplex method developed in Section 7.1 and analyzed in detail in Section 8, we also tested the *Montage* code that uses an iterative algorithm to solve either scalar or planar sky offsets. Figure 17(a) shows the surface brightness difference between our simplex solution and the iterative *Montage* mosaic solution assuming scalar sky offsets. Despite an identical data set, the two methods yield systematic differences of up to ∼0.5 mag arcsec^{-2} at 20 kpc, though the solutions are consistent in the inner disk. Although the simplex and *Montage* scalar-offset mosaics *appear* comparable to the eye, a unique and optimal sky offset solution either does not exist or is extremely difficult for our optimization algorithms to find.

Montage is also capable of fitting planar sky offsets to images, which is a tempting solution to the field-to-field discontinuities that persist between scalar-offset blocks. The result of planar fitting is shown in Figure 13(c). We see that planar sky offsets, in this case, do little to improve the mosaics and indeed have a dramatic effect on the systematic surface brightness of the mosaic (by more than 1 mag arcsec^{-2} in the K_s band). Since our uncertainty on the shape of individual frames (0.3% of the background level across a WIRCam frame) is significant compared to the uncertainty in background level, fitting planar offsets to each WIRCam field introduces additional systematic error propagation compared to scalar sky offsets. *We thus recommend against using planar, or higher-order, sky offsets in wide-field WIRCam mosaics.*

9.2. Comparison to Spitzer/IRAC Images

We also explore systematic uncertainties in our WIRCam mosaics with comparisons against well-calibrated images of

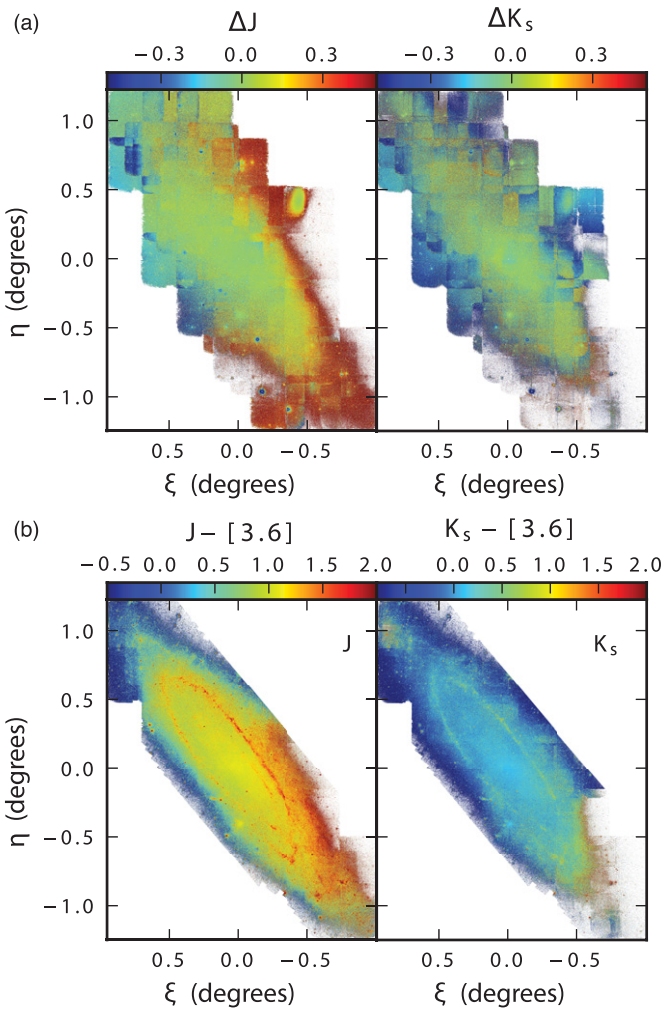

Figure 17. Surface brightness difference maps, showing systematic uncertainties in surface brightness reconstruction. (a) Difference maps between our simplex scalar-fit mosaics (Figure 13(b)) and *Montage* scalar-fit mosaics. (b) Maps of $J - [3.6]$ and $K_s - [3.6]$ surface color inferred from the simplex scalar-sky fitted WIRCam mosaics and *Spitzer*/IRAC 3.6 μm image (Barmby et al. 2006). Note that the IRAC map crops the ANDROIDS/WIRCam footprint.

(A color version of this figure is available in the online journal.)

M31. A template for the NIR disk is the 3.6 μm *Spitzer*/IRAC map, presented in Barmby et al. (2006). Note that although *Spitzer*data avoid background subtraction issues caused by the NIR sky, planar sky offsets were used by Barmby et al., though presumably of a smaller magnitude than our WIRCam sky offsets. In Figure 17(b), we compare our simplex scalar-fitted mosaics against the 3.6 μm image. Generally the $J - [3.6]$ and $K_s - [3.6]$ colors decrease with disk radius, but increase in the star-forming regions due to hot dust emission. However, both color maps (coincidentally) become redder in the southwestern disk beyond the 10 kpc star-forming ring. We interpret this as a systematic over-subtraction of background in these regions on the order of $\gtrsim$1 mag arcsec^{-2}. Evidently, our scalar sky offset mosaics are not systematically reliable beyond the bright disk of M31 with $R > 15$ kpc.

9.3. Monte Carlo Analysis of Systematic Surface Brightness Uncertainties

The difference images presented in the previous section illustrate how the surface brightness reconstructions of identical

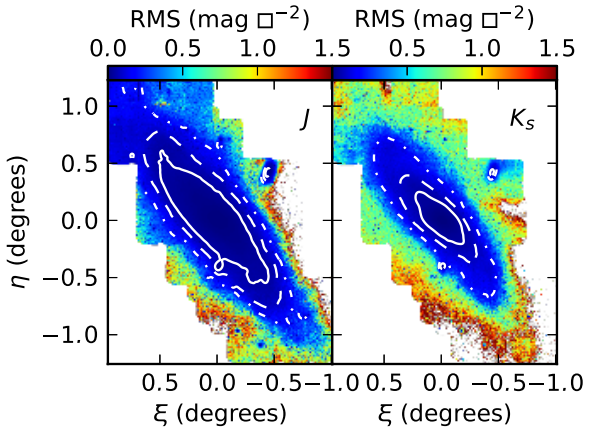

Figure 18. Mosaic maps of bootstrap rms surface brightness in J (left) and K_s (right). White contours identify rms levels of 0.05 (solid), 0.1 (dashed), and 0.2 (dash-dotted) mag arcsec^{-2}.

(A color version of this figure is available in the online journal.)

data can vary depending on the optimization algorithm. Here we pose a slightly different question: how are reconstructions affected by the initial conditions of background errors? That is, given the possible sets of background level biases affecting the blocks, what is the distribution of surface brightness reconstructions? We answer this with a realistic Monte Carlo (MC) analysis.

Our MC realization is generated by perturbing the surface brightness of the corrected blocks with a background error drawn (with replacement) from the ensemble of block sky offsets observed in the original mosaic (Figure 13(b)). Using the scalar-sky fitting procedure, sky offsets are optimized against the known sky background perturbations; 100 such realizations are made to compile an ensemble of mosaics in both bands. Figure 18 shows the rms deviation of MC mosaic surface brightness against the original scalar-fitted mosaics. Reconstructed surface brightness in the outer disk can vary by $\sim$1 mag arcsec^{-2}, consistent with color biases in the $J-[3.6]$ and $K_s-[3.6]$ maps.

We can ultimately understand the source of these systematic surface brightness errors by examining the standard deviations in the residual between expected and realized sky offsets in each MC iteration. This residual dispersion is 0.15% of the J background (0.17% of the K_s background); we find this dispersion to be constant across all fields in the mosaics. If mosaic surface brightness uncertainty is caused by flexure in the mosaic—where blocks on the mosaic periphery are forced to conform to the surface brightness of more central and tightly coupled blocks—then outer blocks would have higher offset dispersion. This is not the case.

Rather than mosaic flexure, a better model for Figure 18 involves uncertainties in the *post priori* adjustment for zero net offset (Equation (5)). Since block sky offsets have approximately Gaussian distributions with dispersions given in Table 2, the uncertainty in the net offset correction is simply $\sigma(\mathrm{block})/\sqrt{n_{\mathrm{blocks}}}$, where $n_{\mathrm{blocks}} = 39$ in the combined 2007B and 2009B mosaic. Given that $\sigma_{\Delta_B} \sim 1\%$, the expected uncertainty in the net offset correction is 0.16%—in perfect correspondence to the observed mosaic uncertainty. The dominant source of uncertainty shown in the MC simulations, Figure 18, is the use of an arithmetic mean of offsets to set an absolute zero point, not flexure or uncertainty in the network of offsets. This suggests that external zero points could be very useful in replacing Equation (5). Since no absolutely calibrated NIR photometry of M31's surface brightness exists, we will discuss a method using panchromatic resolved stellar populations in a future work.

10. CONCLUSIONS

We have presented NIR (J and K_s) images of M31's entire bulge and disk with CFHT/WIRCam. These maps surpass the 2MASS (Beaton et al. 2007) and *Spitzer* (Barmby et al. 2006) mosaics with superior resolution that permits the identification of individual stars throughout M31's mid- and outer disk. The data set is also complementary to the *HST*/WF3 PHAT survey (Dalcanton et al. 2012) by providing complete coverage of M31's entire disk within $R = 22$ kpc, and by offering a broader NIR color baseline ($J - K_s$) than is offered by WF3 (approximately $J-H$). NIR mosaics of M31 have crucial applications for studies of the nearly attenuation-free stellar structure of our nearest spiral neighbor and for tests of stellar population synthesis models in NIR regimes.

Our focus in this paper has been the establishment of procedures for accurately recovering the NIR surface brightness across 3 deg^2 of the M31 disk using an ST nodding observing strategy with WIRCam on CFHT. We have compared two different observing methods to study the effects of ST nodding cadences and patterns on sky subtraction uncertainties. We have also developed and tested flat fielding, zero-point estimation, median sky subtraction, and sky offset optimization procedures in our WIRCam pipeline.

The surface brightness accuracy of a WIRCam frame is affected by both flat-field uncertainties and additive background uncertainties. We recommend using sky flats built from images collected every night to calibrate WIRCam images since these capture the gain structure of WIRCam detectors (unlike dome flats). We also tested sky flats built on longer time spans (across a WIRCam queue run) or shorter spans (updated every half hour), but find that these either introduce flat-fielding biases or become responsive only to changes in additive background contamination, respectively. Although an additive background (e.g., thermal or scattered light) contaminates these flats, the influence of flat-fielding errors on surface brightness shapes is minimal across the background-dominated mosaic. Instead, we find that the surface brightness across WIRCam frames is uncertain by 0.3% of the background intensity due to variations in the background between sky and target fields.

The necessity of nodding between sky and target fields limits our direct knowledge of the background level on the disk by $\gtrsim$2% of the background level. Strictly minimizing latency between sky and disk integrations (as in the 2009B program) provides a 30% reduction in sky offsets, but is ultimately limited by overheads in nodding the telescope and spatial structure in the NIR skyglow itself. Sky offset optimization is successful in reducing block-to-block surface brightness differences to <0.1% of the background level. Our optimization algorithm reliably finds a consistent sky offset solution, so any errors in surface brightness shape across the mosaic are caused by errors in the shapes of individual blocks. There is, however, an uncertainty in the zero point of sky offsets, of order $\sim\sigma_{\Delta_B}/\sqrt{N_{\mathrm{blocks}}}$, 0.16% of the background level. This zero point will ultimately be established using resolved star counts, the subject of a forthcoming ANDROIDS study.

Our experience suggests that wide-field NIR programs that require large ST nods must tune their observing strategies with

sky offset optimization in mind, besides simply minimizing sky sampling latency. Sky offset optimization is aided by having many independent blocks covering the target to decrease the statistical zero-point uncertainty. Increasing the number of independent blocks (observed hours or even a night apart to decouple sky and instrumental biases) is the most reliable way to establish the absolute surface brightness accuracy of the mosaic. Since sky offsets are further biased by any surface brightness shape errors in blocks (realized as our inability to diminish block-to-block offsets below $\sim$0.1% of background brightness), we propose that blocks be interlaced by 50% (so that one detector completely overlaps a detector from an adjacent block). This interlacing pattern would enable the marginalization of shape errors across the entire detector frame. By doubling the number of blocks, each with individually halved exposure times, the mosaic could be reproduced with an equivalent net integration time.

In future work we shall use these NIR mosaics to characterize the structure, stellar mass distribution, and stellar populations of M31 in conjunction with optical CFHT/MegaCam imaging also acquired by the ANDROIDS survey (Sick et al. 2013). We are also compiling a panchromatic catalog of resolved stars across the M31 disk. By directly identifying the stars that contribute to the NIR light of a galaxy, we will assess long-standing tensions between NIR isochrones and broadband SED modeling of galaxies.

We thank the anonymous referee for offering many useful suggestions that clarified this work. We also thank Loic Albert and Karun Thanjavur, previously from CFHT, for their help with WIRCam I'iwi data products and procedures and the CFHT staff for many informative discussions and their diligence in performing the queue service observations. We are grateful for the *Spitzer* 3.6 μm mapping provided by Pauline Barmby (University of Western Ontario). J.S. and S.C. acknowledge support through respective Graduate Scholarship and Discovery grants from the Natural Sciences and Engineering Research Council of Canada. M.M. is supported by NASA through a Hubble Fellowship grant HST-HF51308.01-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. This work is based on observations obtained with WIRCam, a joint project of Canada–France–Hawaii Telescope (CFHT), Taiwan, Korea, Canada, and France, at the CFHT, which is operated by the National Research Council (NRC) of Canada, the Institute National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii.

Facilities: CFHT (WIRCam), FLWO:2MASS

REFERENCES

Adams, J. D., & Skrutskie, M. F. 1996, Airglow and 2MASS Survey Strategy, http://astsun.astro.virginia.edu/~mfs4n/2mass/airglow/airglow.html
Athanassoula, E., & Beaton, R. L. 2006, MNRAS, 370, 1499
Barmby, P., Ashby, M. L. N., Bianchi, L., et al. 2006, ApJL, 650, L45
Beaton, R. L., Majewski, S. R., Guhathakurta, P., et al. 2007, ApJL, 658, L91
Berriman, G. B., Good, J. C., Laity, A. C., & Kong, M. 2008, in ASP Conf. Ser. 394, Astronomical Data Analysis Software and Systems XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis (San Francisco, CA: ASP), 83
Bertin, E. 2006, in ASP Conf. Ser. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique (San Francisco, CA: ASP), 112
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in ASP Conf. Ser. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley (San Francisco, CA: ASP), 228
Brown, T. M., Beaton, R., Chiba, M., et al. 2008, ApJL, 685, L121
Brown, T. M., Ferguson, H. C., Smith, E., et al. 2003, ApJL, 592, L17
Brown, T. M., Smith, E., Ferguson, H. C., et al. 2006, ApJ, 652, 323
Brown, T. M., Smith, E., Ferguson, H. C., et al. 2009, ApJS, 184, 152
Bruzual, G. 2007, in IAU Symp. 241, Stellar Populations as Building Blocks of Galaxies, ed. A. Vazdekis & R. Peletier (Cambridge: Cambridge Univ. Press), 125
Conroy, C. 2013, ARA&A, 51, 393
Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833
Courteau, S., Cappellari, M., de Jong, R. S., et al. 2014, RvMP, 86, 47
Courteau, S., Widrow, L. M., McDonald, M., et al. 2011, ApJ, 739, 20
Dalcanton, J. J., Williams, B. F., Lang, D., et al. 2012, ApJS, 200, 18
de Jong, R. S. 1996, A&A, 313, 377
Dutton, A. A., Courteau, S., de Jong, R., & Carignan, C. 2005, ApJ, 619, 218
Ibata, R., Chapman, S., Ferguson, A. M. N., et al. 2005, ApJ, 634, 287
Irwin, M. J., Ferguson, A. M. N., Ibata, R. A., Lewis, G. F., & Tanvir, N. R. 2005, ApJL, 628, L105
Kormendy, J., & Kennicutt, R. C., Jr. 2004, ARA&A, 42, 603
MacArthur, L. A., Courteau, S., Bell, E., & Holtzman, J. A. 2004, ApJS, 152, 175
Maraston, C. 1998, MNRAS, 300, 872
Maraston, C. 2005, MNRAS, 362, 799
Maraston, C., Daddi, E., Renzini, A., et al. 2006, ApJ, 652, 85
Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883
Marmo, C., & Bertin, E. 2008, in ASP Conf. Ser. 394, Astronomical Data Analysis Software and Systems XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis (San Francisco, CA: ASP), 619
Massey, P., Olsen, K. A. G., Hodge, P. W., et al. 2006, AJ, 131, 2478
McConnachie, A. W., Irwin, M. J., Ferguson, A. M. N., et al. 2005, MNRAS, 356, 979
McConnachie, A. W., Irwin, M. J., Ibata, R. A., et al. 2009, Natur, 461, 66
Nelder, J. A., & Mead, R. 1965, CompJ, 7, 308
Olsen, K. A. G., Blum, R. D., Stephens, A. W., et al. 2006, AJ, 132, 271
Pforr, J., Maraston, C., & Tonini, C. 2012, MNRAS, 422, 3285
Press, W. H., Tuekolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical Recipes: The Art of Scientific Computing (3rd ed.; Cambridge: Cambridge Univ. Press)
Puget, P., Stadler, E., Doyon, R., et al. 2004, Proc. SPIE, 5492, 978
Saglia, R. P., Fabricius, M., Bender, R., et al. 2010, A&A, 509, A61
Sick, J., Courteau, S., & Cuillandre, J.-C. 2013, in Structure and Dynamics of Disk Galaxies, ed. M. S. Seigar & P. Treuthardt (San Francisco, CA: ASP), in press (arXiv:1310.4832)
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, MNRAS, 418, 1587
Vaduvescu, O., & McCall, M. L. 2004, PASP, 116, 640
Williams, B. F. 2002, MNRAS, 331, 293
Williams, B. F. 2003, AJ, 126, 1312
Worthey, G., España, A., MacArthur, L. A., & Courteau, S. 2005, ApJ, 631, 820