Some Thoughts on Cross-calibration in the Mid-Infrared

Sean J. Carey
Spitzer Science Center, IPAC, Caltech, Pasadena, CA, 91125

Abstract. I provide some examples of the cross-calibration between the instruments aboard the Spitzer Space Telescope. Current cross-calibration accuracy between the Spitzer instruments is better than 5% and consistent with the calibration uncertainties of the instruments and the current uncertainties in the absolute calibration of infrared zero-points. Utility of cross-calibration extends beyond reconciling measured fluxes; examples are given of how cross-calibration can be used to identify/mitigate instrumental signatures. The Spitzer cross-calibration efforts are placed in context by a brief review of the current state of absolute calibration in the infrared. The current accuracy of calibration is limited by knowledge of the fundamental zero points to about 2%. Rapid progress in the field is being made which suggests that the 1% absolute photometric accuracy requirement of some future dark energy probes can be met.

1. Introduction

Cross-calibration, or comparison to the absolute photometric accuracy of other observatories/instruments is fundamental to placing data from one instrument in scientific context with existing data. Note that for the remainder of this contribution, I will use instrument as the generic term for a fundamental quanta of device that conducts observations that need to be calibrated. In addition to verifying that observations from different instruments can be placed on the same physical scale, cross-calibration is often used during development and data reduction to identify other calibration issues which can be missed due to the resource-limited nature of most space-based astronomical instruments. As a general rule, the best way to improve and understand the calibration of any individual data set is to use it for science which necessarily means comparing the data to other existing data sets for most (but not all applications). Section 3 provides specific examples of cross-calibration between the various instruments aboard the Spitzer Space Telescope (Werner et al. 2004). These examples span the gamut of uses from refining the absolute calibration, to correcting for instrumental signatures to sanity checks for science purposes.

Cross calibration is related to but is often a separate activity from the absolute calibration of a instrument. Absolute calibration is the tying of the observed (spectro)photometry of an instrument to a known physical standard. Ideally, there would be no need for cross-calibration if all instruments were perfectly absolutely calibrated using the same or consistent fundamental standards. However, this is particularly not the case in the infrared as different authors use different standards or different values for the same standards. As a result, the differences arising in the cross-calibration of instruments can be attributed to uncertainties/errors in the instrument calibrations and differences in the assumed fundamental calibrator used in the calibration of each instrument. Section 2 provides a summary of the current state of absolute calibration in the infrared to place the cross calibration examples discussed in this paper in better context.

Fortunately, for many science goals, a perfect absolute calibration is not essential. In fact, for exoplanet transit observations, arguably one of the more exciting applications of
space-based infrared observations from Spitzer, the relative calibration is also unimportant as long as the calibration remains stable throughout the observations. Implicit in this statement is the assumption that the parameters of the host star (effective temperature and radius) are extremely well-known (e.g. van Belle & van Braun 2009), but this detailed characterization is well beyond the scope of most calibration campaigns. It is worth mentioning that as the precision of transiting planet observations improves the knowledge of the fundamental properties of the host stars will be the limiting factor in measuring planet radii and determining their temperatures.

In many cases, as long as the instruments are consistently calibrated then colors and flux ratios will be correct. Derived quantities such as temperature, hardness of incident radiation field or extinction can be well determined from data with a good relative calibration. If the uncertainties in absolute calibration are a few percent or so, which is currently the state of the best infrared calibration, then quantities such as derived luminosity and distance will have small errors that are acceptable for most science applications.

However, absolute calibration is significant for one major line of inquiry, using supernovae as a probe of dark energy requires absolute photometric accuracies of order 1%. The use of supernovae as probes of dark energy is one of the goals of the WFIRST mission, a primary recommendation of the 2010 Decadal Survey of Astronomy and Astrophysics. Kim et al. (2004) provide a discussion of the effect of zero point shifts on the derivation of cosmological parameters using supernovae as standard candles. Currently, the required absolute calibration is not met in the near- to mid-infrared, but current work on absolute calibration is promising and suggests that this goal could be met in the near future.

In addition to the discussion on absolute calibration in Section 2 and the cross-calibration examples in Section 3, I contemplate a suggestion to improve the absolute and cross-calibration of future space-based instruments in Section 4 and provide some final comments in Section 5.

2. Infrared Absolute Calibration

The absolute calibration of modern infrared instruments is complicated by the high sensitivity of the instruments which precludes direct observation of the fundamental standards. The calibration is typically done by observing a set of standards which are well spectral typed and have excellent photometry at shorter (optical) wavelengths. The absolute flux scaling of the standards is done by transferring the standard to a fundamental calibration through optical and near-infrared photometry. A variety of standards are used: A dwarfs as they are relatively featureless in the infrared, K giants as they are bright in the infrared and solar analogs as there exist good models based on the solar spectrum. White dwarfs have been used extensively in the near-infrared calibration of HST with internal consistencies of 1% (e.g. Bohlin et al. 2001). KIII and solar analog models have uncertainties of order 1-2% at the relevant wavelengths. Figure 1 displays the difference in the Cohen et al. (2003) and Engelke et al (2006) templates for the KIII star, NPM1p68.0422, used as a calibrator for IRAC (Fazio et al. 2004). The differences in the templates are profound at the several percent level and attest to the current difficulties in mid-infrared calibration. As instruments become increasingly more sensitive and can no longer access the brighter calibrators, solar analogs and white dwarfs are becoming more important and will be key components of the calibration methodology for JWST.

Unfortunately, the fundamental standard, Vega, is problematic. At wavelengths of 12 µm and longer, Vega has a pronounced infrared excess (Aumann et al. 1984) due to circumstellar material. An inner, hotter circumstellar disk (Ciradi et al. 2001) further complicates analysis in the mid-infrared. For shorter wavelengths, it has been shown that Vega is a rapid rotator (e.g. Gray 1988) and is not well represented by models of A0V type stars. Coupled with the uncertainties in the templates/models of the primary standards, the
uncertainties in using Vega as a fundamental calibrator result in the best current infrared calibrations having uncertainties of order 3%. This is reflected by the current state of the literature as different workers in the field can arrive at calibrations that differ by 2% starting with the same data and apparently (at least to this author) reducing and analyzing the data in equally valid ways. This is clearly the case in comparing the work of Rieke et al. (2008) and Price et al. (2004), both of whom heavily weight the MSX data to determine the flux of Vega at around 10 \( \mu m \). Engelke et al. (2010) have recently attempted to improve the situation by developing a fundamental zero point spectrum based on 109 Vir in the optical and Sirius in the infrared. The situation is even worse at longer wavelengths as there are few stars that can be observed and many of the calibrators used such as planets or asteroids are inherently variable.

For Spitzer, each of the three science instruments used a different methodology to develop their primary calibration standards. The choice of calibration method was left to the teams building the instruments and adopted by the Spitzer Science Center during standard operations. For the 3-8 \( \mu m \) cameras comprising the IRAC instrument, a network of four AV and seven KIII standards were developed using absolutely calibrated spectral templates (Cohen et al. 2003). These templates are rooted in the fundamental calibrations associated with MSX and the methodology described in Price et al. (2004). During operations, it was noted that the KIII standards and AV standards produced calibrations that were discrepant by 7.3%, 6.5%, 3.6% and 2.1% for the 3.6, 4.5, 5.8 and 8.0 \( \mu m \) channels, respectively. Reach et al. (2005) adopted the AV star calibration as it was more likely that uncertainties in molecular absorption features in the KIII spectra were complicating the templates. This hypothesis was confirmed through IRTF observations of some of the IRAC primary calibrators (Figure 2). In the final processing of IRAC data, updated templates (see Figure 1) for the KIII stars are being used. With the new templates, the discrepancy between the calibration factor determined using the KIII stars and AV stars is less than 1% in all

Figure 1: Difference in the spectral templates from Cohen et al. (2003) and Engelke et al. (2006) for IRAC K2III calibrator, NPM1p68.0422. Most of the difference can be attributed to the processing of the ISO SWS spectra used to template this spectral type.
Figure 2: Comparisons of spectral templates to observed data. The black curve is the ratio of IRTF data of the K2III star, NPM1p68.0422 to the A1V star, HD165459. The bright A0V star, HD172728, was used as a ratioing standard. The red curve is the ratio of the Cohen et al. (2003) template for NPM1p68.0422 to the Cohen template for HD165459. The blue curve is the ratio of the Engelke et al. (2006) template for NPM1p68.0422 to the Cohen template for HD165459. The IRAC 3.6 µm response function is plotted as a dashed line. The Engelke template is in better agreement with the observations.

channels. The 24, 70 and 160 µm photometers (and 70 µm spectrometry) comprising the MIPS (Rieke et al. 2004) instrument used a combination of AV stars at 24 µm (Engelbracht et al. 2007) and 70 µm (Gordon et al. 2007, Lu et al. 2008 for the spectroscopic mode) and asteroids and red extragalactic sources at 160 µm (Stansberry et al. 2007). I will concentrate on the 24 µm calibration in the remainder of the discussion as that is the MIPS channel that overlaps with other well-calibrated datasets from Spitzer and MSX. The 24 µm calibration uses [K] – [24] = 0 for AV stars and a photospheric model of Vega for the zero point flux density at 24 µm (Rieke et al. 2008). The IRS spectrographs (Houck et al. 2004) are calibrated using MARCS stellar models (Decin et al. 2004) of one primary standard, HR 7341 (K1III spectral type) and several secondary standards.

3. Specific examples of cross-calibration

My case studies in the utility of cross-calibration are all drawn from personal experience and deal with the three instruments, IRAC, MIPS and IRS aboard Spitzer. The examples span the spectrum of uses of cross-calibration from the traditional comparison to diagnosis of issues in the calibration of a specific instrument than cannot be well-handled without the use of data from another instrument. In all cases, the cross-calibration is between instruments that have overlapping bandpasses/spectral responses. Cross-calibration can also be done between instruments without overlapping responses; however, the assumed spectral energy distribution of the calibration sources can become a larger source of error.
for cross-calibration without overlapping responses. For photometers, if the instruments are perfectly, relatively calibrated then

\[ \frac{F_A \times S(\nu_B)}{C_A} = \frac{F_B \times S(\nu_A)}{C_B} \]

(1)

where \( F \) is the measured flux density of instrument A or B, \( C \) is the appropriate color correction for the given instrument and calibration source spectrum and \( S(\nu) \) is the source flux density at the effective frequency of the given instrument. Typically, color-corrections are the order of a few percent for well-behaved calibration sources and for instruments calibrated using the isophotal assumption. The ratio, \( S(\nu_B)/S(\nu_A) \), is will typically differ from 1 by \( \delta \nu/\nu \) for stellar calibrators and similar instrument bandpasses. For extended sources, the variation in effective flux density can be quite large (up to 50%) depending on the spectral energy distribution of the source.

In addition to the examples in Sections 3.1-3.6, ongoing programs exist to cross-calibrate the Spitzer instruments with WISE, IRAC and MIPS with HST and JWST (Gordon & Bohlin, this proceeding) and warm IRAC observations are planned to cross-calibrate with DIRBE. The DIRBE cross-calibration observations will include at least one map of a DIRBE calibrator using IRAC for the entire DIRBE beam (42 arcminutes) to assess how the additional flux in the DIRBE beam due to extended sources and field stars affects the comparison.

### 3.1. Spitzer Instrument cross-calibration

While there is no cross-calibration requirement for Spitzer, the Spitzer Science Center conducted several cross-calibration experiments. The experiments consisted of observations of calibrators for the spectrometer, IRS, by all three instruments as the photometers could achieve high signal-to-noise data with very short observations of the IRS standards. Gizis et al. (2010) will discuss these experiments in great detail, but the salient points are summarized below.

Between IRAC and IRS, observations of three IRS calibrators, HR 7341, HR 2194 (A0V), and HR 6606 (G9III) were compared by integrating the IRS SL1 and SL2 spectra with the IRAC 8 μm response function. For all three sources the photometry agrees to 0.3%. The uncertainty in absolute calibration of IRAC is 3% and IRS has a uncertainty of 5%. Considering that the calibration methodology of the two instruments is entirely different this agreement is surprisingly good. The comparison of IRS to MIPS also uses overlapping bandpasses. In this case, IRS LL observations of HR 2194 and HR 6348 (K1III) are compared to the MIPS 24 μm passband. IRS is 2.2%±1.0% fainter than MIPS at 24 μm. The calibration uncertainty of MIPS is 4%, and while the IRS/MIPS offset is measurable, it is comfortably within the instrumental calibration error budgets. The disagreement is in accord with the good agreement of IRS with IRAC and the results of Rieke et al. (2008) which suggest that IRAC is 1.5% lower than MIPS. The difference between IRAC and MIPS is attributed to the differences in determination of the flux density of Vega at 10 μm between Rieke and MSX (Price et al. 2004) as the IRAC templates developed by Cohen are based on the MSX results.

### 3.2. IRAC extended source

Comparison between different instruments can often help characterize instrumental signatures/artifacts that would be difficult to diagnosis without external information. One example is the pronounced internal scattering present in the Si:As detectors used by IRAC in the 5.8 and 8.0 μm passbands. Approximately 30% of the light incident on an IRAC Si:As detector pixel is scattered more or less uniformly throughout the remainder of the array. Cohen et al. (2007) independently confirmed this result. As IRAC is calibrated using point sources and a small photometric aperture (12.2 arcsecond radius), this scattered light
3.3. Warm IRAC / Cryogenic IRAC

Cross-calibration is also useful for the same instrument if the operating parameters change significantly. One such example is the operation of the IRAC InSb detector arrays at 28.7 K after the exhaustion of cryogen ending the prime mission for Spitzer. As part of the warm recalibration, the photometry of the primary IRAC calibrators was compared to the cryogenic values. Not surprisingly, the flux conversions between data numbers and
flux density changed as the detectors operated at significantly higher temperatures (during cryogenic operations the detectors were at 15 K) and at a lower applied bias setting (500 mV compared to 750 mV) for the 3.6 µm array. However, we were also able to improve the calibration of the cryogenic data with new information on the nature of the intra-pixel gain variations exhibited in these detectors. In warm operations, the variations are 2× and 4× more significant at 3.6 and 4.5 µm, respectively. Figure 4 displays an intra-pixel gain map made from dedicated observations during the checkout phase of the warm mission. The gain variations for each pixel are parameterized by a two-dimensional Gaussian in x and y offset of the centroid of a star from the center of a pixel.

As the significance of the intra-pixel gain variation was not understood during the initial characterization phase of the cryogenic mission, no dedicated experiment was conducted to determine the nature of the effect. In the Reach et al. (2005) calibration, a one-dimensional parameterization was used which left a residual scatter of ∼1%. With the additional information from the warm characterization, the final cryogenic processing was able to use the same 2d Gaussian parameterization and significantly reduce the scatter in the resulting photometry.
3.4. MIPSGAL point source

The MIPSGAL legacy survey (Carey et al. 2009) mapped the inner Galactic plane with MIPS at 24 and 70 $\mu$m. It is complementary to the GLIMPSE survey (Churchwell et al. 2009) which covered the same region of the Milky Way with IRAC. To extract science from the combined data sets, the relative accuracy of the photometry of the data products needed to be assessed. As part of the preparation of the MIPSGAL point source catalog, we compared the measured [8] – [24] color for a sample of KIII and AV stars imaged by both surveys. Figure 6 displays the difference in observed color from the predicted color. Several of the stars (about 15%) in the sample had significantly redder color than predicted which is indicative of circumstellar disks. After removing the disk candidates (color differences less than -0.2 magnitude), the average color difference is +0.03 indicating that the IRAC 8 $\mu$m photometry is 3% too bright compared to the 24 $\mu$m photometry which is contrary to the results of Section 2 (although still within the uncertainties) and has also been noted by the SAGE legacy team in their data of the Large Magellanic Cloud. This result is inconsistent with the cross-calibration of Section 3.1 and may be possibly due to the different methods used to determine the photometry.

3.5. MIPSGAL extended source

As part of the MIPSGAL processing, a more exact droop correction was applied in the presence of saturated data. This and other pipeline enhancements specific to the bright, diffuse backgrounds of the Galactic plane are discussed in Mizuno et al. (2008). To verify that these enhancements did not adversely effect the data, we smoothed the 24 $\mu$m image data to the same resolution as the MSX 21.3 $\mu$m data and compared the surface brightness on a per-resolution element basis. Assuming that underlying emission had a spectrum resembling a massive star-forming region such as M16, we color corrected each data set and compared the data to a $\nu^{-1.5}$ spectrum. The MSX and MIPS data are in excellent agreement up to about 1700 MJy sr$^{-1}$ at which point the MIPS detectors saturate.
Some Thoughts on Cross-calibration in the Mid-Infrared

Figure 7: Comparison of surface brightness between MIPS 24 μm and MSX 21.3 μm. The greyscale is density of points from 3 to 100 resolution elements per 1 MJy sr⁻¹ × 1 MJy sr⁻¹ bin. Crosses are individual pixels outside of the lowest contour. The line is the correlation expected for infrared-bright regions of the ISM.

3.6. MIPSGAL/HiGal

At wavelengths longer than 24 μm, absolute calibration is extremely challenging. As an independent check of the photometry of the Herschel Galactic plane survey (HiGal; Molinari et al. 2010), the HiGal team compared the 70 μm PACS maps to the 70 μm MIPSGAL maps of the same regions. The result for one of the 2°× 2° tiles is shown in Figure 8. The agreement is quite good considering the degree of striping still present in the MIPS 70 μm maps and the preliminary state of the PACS calibration.

4. A Modest Calibration Proposal

With an increasing number of observatories planned for L2, a useful resource for those instruments would be an NIST referenced, external calibrator similar in spirit to the aluminum reference spheres used by MSX (Price et al. 2004). The idea would be to fly a reference source which is absolutely calibrated to better than 1%. Unlike the MSX reference spheres which were passive in nature, the calibrator should be active and able to adjust its temperature and output luminosity. The calibrator should have an internal integrating sphere with some ability to adjust the spectral signature either through filters at the output window of the cavity or perhaps a Fabry-Perot to output a pure spectral frequency. By adjusting the output, the calibrator would serve instruments from the UV to the submillimeter. The calibrator should provide telemetry on power, temperature and location. The power would be supplied via solar panels probably requiring the calibration source to be either on a boom or tethered. Consumables would consist of propellant for station keeping at L2. The calibration source would probably not require its own launch vehicle; it is likely that it could be launched along with a larger mission that had some space in the fairing.
My back of the envelope estimate for cost is $\approx 50$ million which is likely to be $<1\%$ of the cost of missions sent to L2.

5. Concluding Remarks

While the current state of the absolute calibration in the infrared is not optimal, there has been much recent progress and a good reason to be optimistic that the calibration will achieve the better than 1% requirement needed for WFIRST. An improved absolute infrared calibration will also greatly facilitate science to be done with JWST although it is not needed to achieve any of the major science goals. There has been excellent recent progress in attempting to move away from Vega as a primary calibrator (Engelke et al. 2010), improve stellar models (Castelli & Kurucz 2004; Decin & Eriksson 2007), and unify the calibration scheme through well-planned cross-calibration experiments such as the HST/Spitzer work of Bohlin & Gordon (this proceeding, also Bohlin et al. 2010). The ACCESS mission (Kaiser, this proceeding), a rocket borne, NIST standard calibrated mission will help provide a framework for securing the photometry of stellar standards to an absolute scale. Ongoing development of calibration sources for instruments such as SNDICE (Schahmaneche, this proceeding) and ground based absolute calibration campaigns such as described by Zimmer et al. (2010) will continue to improve the situation. The calibration mission sketched in Section 4 would permit absolute flux measurements to be made limited only by the signal-to-noise of the science observation and would obviate the need for cross-calibration.

Acknowledgments. The author is grateful for the efforts of the combined Spitzer team and is most indebted to the efforts of the IRAC instrument support team. I have benefitted from many useful discussions with Stephan Price, John Gizis, Tom Jarrett, Jason Surace, Mark Lacy and Bill Reach.
References

Bohlin, R. C. et al. 2010, in preparation
Gizis, J., in preparation
Gray, R. O. 1988, JRASC, 82, 336