PARALLELIZATION AND ALGORITHMIC ENHANCEMENTS OF HIGH RESOLUTION *IRAS* IMAGE CONSTRUCTION

Yu Cao, Thomas A. Prince

Department of Physics, California Institute of Technology, Pasadena, CA 91125

Susan Terebey, Chas Beichman

Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125

August 1995

ABSTRACT

The *Infrared Astronomical Satellite* carried out a nearly complete survey of the infrared sky, and the survey data are important for the study of many astrophysical phenomena. However, many data sets at other wavelengths have higher resolutions than that of the co-added *IRAS* maps, and high resolution *IRAS* images are strongly desired both for their own information content and their usefulness in correlation studies.

The HIRES program was developed by the Infrared Processing and Analysis Center (IPAC) to produce high resolution (˜ $1'$) images from *IRAS* data using the Maximum Correlation Method (MCM). We describe the port of HIRES to the Intel Paragon, a massively parallel supercomputer, other software developments for mass production of HIRES images, and the *IRAS* Galaxy Atlas, a project to map the Galactic plane at 60 and 100 μm.

Images produced from the MCM algorithm sometimes suffer from visible striping and ringing artifacts. Correcting detector gain offsets and using a Burg entropy metric in the reconstruction scheme were found to be effective in suppressing these artifacts. A variation of the destriping algorithm was used to subtract zodiacal emission.

1. INTRODUCTION

The *Infrared Astronomical Satellite* (*IRAS*) provided our first comprehensive look at the infrared sky, producing a nearly complete survey at mid- to far-infrared wavelengths (12, 25, 60, and 100 μm) (Beichman (1987); Soifer, Houck, and Neugebauer (1987); *IRAS* Catalogs and Atlases: Explanatory Supplement (1988)). Images made from the *IRAS* survey data show a wealth of extended structure from star-forming regions and other components of the interstellar medium. A variety of studies exploiting the *IRAS* images have been made to date ranging from structure on a galactic scale to detailed studies of individual molecular clouds (e.g. Beichman et al. (1986); Weiland et al. (1986); Terebey and Fich (1986); Boulanger and Perault (1988); Sodroski et al. (1989); Scoville and Good (1989); Snell, Heyer, and Schloerb (1989); Clemens, Yun, and Heyer (1991); Wood et al. (1994)). The strength of *IRAS* is the completeness of the survey. However, in many cases the spatial resolution of the comparison data sets at other wavelengths is better than for *IRAS*, and thus the $4' - 5'$ resolution of the released *IRAS* images (the Infrared Sky Survey Atlas, ISSA, Wheelock et al. (1994)) can limit the comparison. The desire for higher spatial resolution combined with the paucity of new infrared satellite missions has inspired many efforts to extract high spatial resolution information from the data (e.g. Bontekoe et al. (1994);). The products most widely accessible to the US science community are the HIRES images distributed by the Infrared Processing and Analysis Center (IPAC), which are based on the Maximum Correlation Method (MCM) (Aumann, Fowler, and Melnyk 1990). The HIRES images have been successfully used for a variety of galactic and extragalactic studies (Rice (1993); Surace et al. (1993); Terebey and Mazzarella (1994)).

Application of the HIRES algorithm to the *IRAS* data has been limited largely by the computational resources available for HIRES processing. A $1° \times 1°$ field of typical scan coverage takes 1 hour of CPU time on a Sun SPARCstation 10, for all four wavelength bands and 20 iterations (at which point artifacts limit further improvement of image quality). To overcome these CPU limitations we have undertaken the porting of the HIRES software to the Intel Delta and Paragon parallel supercomputers. HIRES processing is now feasible for large regions of the sky.

As part of a program in high-performance computational science and engineering, Caltech has developed significant software and hardware capabilities for massively parallel computing (also called *concurrent supercomputing*). Among the several concurrent computers currently available at Caltech is the 512-node Intel Touchstone Delta, a prototype parallel supercomputer with measured performance of 13 GFLOPS, 8 Gigabytes of memory, and 90 Gigabytes of disk. Upgraded resources include a 56-node and a 512-node Intel Paragon. The new 512-node Intel Paragon Model L38, has a peak speed of 38.4 GFLOPS, 16 Gigabytes of memory, and 14 RAIDs that control 67.2 Gigabytes of disk, one Ethernet node, two HIPPI nodes, and six service nodes. The high demand for HIRES images, along with the availability of parallel computing facilities, motivated the port of HIRES to the parallel supercomputers.

The development of new artifact reduction algorithms allows the iterative procedure to be carried much further, requiring more CPU time and further justifying the parallel computing approach.

These efforts made possible a large scale mapping project: high resolution *IRAS* mapping of the Galactic plane. The new *IRAS* Galaxy Atlas (IGA) maps provide a 20-fold improvement in areal information content over current *IRAS* 60 and 100 μm maps and will be valuable for a wide range of scientific studies, including:

- The structure and dynamics of the interstellar medium (ISM)
- Cloud core surveys within giant molecular clouds
- Determination of initial mass functions (IMFs) of massive stars
- Study of supernova remnants (SNRs)

The IGA images will be made available on-line at IPAC. Additional information will come from combining the 60 and 100 μm HIRES data with the images and catalogs being produced from the 12 and 25 μm *IRAS* data by the Air Force Phillips Laboratory and Mission Research Corporation. Alternatively, standard four band HIRES images can be requested from IPAC.

The original HIRES algorithm which produces high resolution *IRAS* images and later enhancements are described in Section 3. The Maximum Correlation Method (MCM) algorithm (Aumann, Fowler, and Melnyk 1990) produces high resolution images from the survey and additional observation (AO) data, using a nonlinear iterative scheme. The resulting images have resolution of about $1'$, compared to the $4' - 5'$ subtended by the 100 μm band detectors in the *IRAS* focal plane. A description of the basic MCM algorithm is outlined in Section 3.1. In Sections 3.2 and 3.4 we offer descriptions of artifact reduction algorithms, namely using estimates of gain offset to eliminate striping, and using a Burg entropy metric in the iterative algorithm to suppress ringing around bright point sources. Detector data calibration and zodiacal light subtraction are carried out with a spin-off method of the destriping algorithm (Section 3.3). Validation of the algorithmic enhancements and output image properties is given in Section 4.

In the parallel processing each $1^\circ \times 1^\circ$ image field is mapped to an 8- or 16-node process grid, which shares the computation by loading different observation scans. An efficiency of 60 % is reached with 8 nodes. The parallelization strategy and pipeline implementation (which coordinates computation and data transfer on workstations and supercomputers) will be discussed in Section 5.

2. RELEVANT INFORMATION ABOUT *IRAS*

The *IRAS* survey was designed for the identification of point sources, rather than as an imaging instrument. The data were taken with rectangular detectors that scanned the sky multiple times in "push-broom" fashion (e.g. see Fig. 2). The satellite data are fundamentally in the form of one-dimensional data streams for each detector. During post-processing it was discovered that two-dimensional images could be made by stitching together, i.e.

focal-plane.eps

Figure 1: The *IRAS* Focal Plane

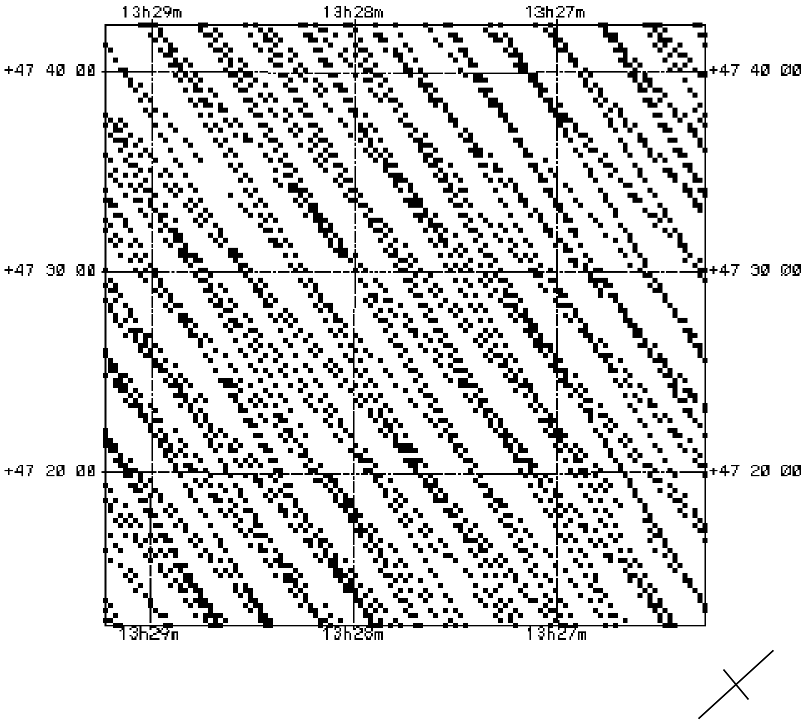

Figure 2: *IRAS* scan pattern in M51. Dots represent 60 μm detector footprint center positions. Lower right cross indicates FWHM of the 60 μm detector response function.

coadding, these one-dimensional detector streams. This basic processing accounts for many of the characteristics of the *IRAS* images. For example, stripes are a common image artifact because there are small gain variations in the one-dimensional detector streams. Also, the shape of the beam varies from place to place because the coverage, meaning number and orientation of one-dimensional detector streams, is nonuniform. The effective data oversampling make the *IRAS* data amenable to resolution enhancement because of the geometric information contained within overlapping data samples.

The *IRAS* focal plane (shown in Fig. 1) included eight staggered linear arrays subtending $30'$ in width, two in each of four spectral bands at 12, 25, 60, and 100 μm. Data rate considerations forced the detector sizes to be much larger than the diffraction limit of the telescope. The typical detector sizes were 45×267, 45×279, 90×285, and 180×303 arcsec (full width at half maximum response, FWHM) respectively, at the four wavelength bands.

This combination of focal plane, detector size, and scan pattern optimized detection of point sources in areas of the sky where the separation between sources was large compared to the sizes of the detectors. However, it complicates the construction of images of regions containing spatial structure on the scale of arcminutes.

3. ALGORITHM

A typical HIRES processing consists of the following steps:

1. Uncompressing and extracting calibrated data scans from archive;
2. Data preprocessing, including cross-scan offset calibration, baseline removal, deglitching, [1] and noise estimation;
3. Subtraction of zodiacal emission (optional);
4. Reprojecting data to desired geometry (optional);

---

[1] Deglitching stands for the removal of spurious non-source-like signals called glitches, typically caused by cosmic ray events in individual detectors.

TABLE 1
Definition of Notations

D_i	measured detector flux at data sample i
f_j	intensity at image pixel j
$f_j^{(k)}$	estimate of f_j at iteration k
$F_i = \sum_j r_{ij} f_j$	mock data flux at sample i given image f_j
$C_i = D_i / F_i$	detector correction factor
c_j	pixel correction factor
G_L	gain offset of detectors in scanline L
$D_i^* = G_L D_i$	gain compensated detector flux
$C_i^* = D_i^* / F_i$	gain compensated detector correction factor
ΔS	Burg entropy difference between two images f_j and $f_j + \Delta f_j$
$g_{jj'}$	Burg entropy metric tensor

5. High resolution image reconstruction.

This section describes the algorithmic aspects of HIRES, specifically, issues involved in steps 2, 3, and 5. For details on the technical aspects of the implementation, see Section 5.

Table 1 summarizes the notations used in this section.

3.1. *The Maximum Correlation Method*

Starting from a model of the sky flux distribution, the HIRES MCM algorithm folds the model through the *IRAS* detector responses, compares the result track-by-track [2] to the observed flux, and calculates corrections to the model. One important characteristic is that the standard MCM algorithm conserves flux. The process is taken through about 20 iterations at which point artifacts limit further improvement. The algorithm yields a resolution of approximately $1'$ at 60 μm. This represents an improvement in resolution by as

[2]Track, also called *leg* or *scanline*, refers to the set of data samples collected consecutively by one detector moving across a given field.

much as a factor of 20 in solid angle over the previous images from the *IRAS* Full Resolution Survey Coadder (FRESCO). We give a brief description of the MCM algorithm following the formalism and notations of Aumann, Fowler, and Melnyk (1990).

Given an image grid f_j, with n pixels $j = 1, ..., n$ and m detector samples (*footprints*) with fluxes $D_i : i = 1, ..., m$, whose centers are contained in the image grid, an image can be constructed iteratively from a zeroth estimate of the image, $f_j^0 = \text{const.} > 0$ for all j. In other words the initial guess is a uniform, flat, and positive definite map. For each footprint, a correction factor C_i is computed as,

$$C_i = D_i/F_i, \tag{1}$$

where

$$F_i = \sum_j r_{ij} f_j, \tag{2}$$

and r_{ij} is the value of the ith footprint's response function at image pixels f_j. Therefore F_i is the current estimate of the ith footprint's flux, given image grid f_j.

A mean correction factor for the jth image pixel is computed by projecting the correction factor for the footprints into the image domain:

$$c_j = [\sum_i (r_{ij}/\sigma_i^2) C_i]/[\sum_i (r_{ij}/\sigma_i^2)]. \tag{3}$$

The weight attached to the ith correction factor for the jth pixel is r_{ij}/σ_i^2, where σ_i is the *a priori* noise assigned to the ith footprint.

The kth estimate of the image is computed by

$$f_j^{(k)} = f_j^{(k-1)} c_j. \tag{4}$$

In practice when the footprint noise σ_i is not easily estimated, an equal noise value for all footprints is assumed, and the MCM is identical to the Richardson-Lucy algorithm (Richardson (1972); Lucy (1974)).

3.2. *Destriping Algorithm*

Stripes are the most prominent artifacts of the HIRES images. HIRES takes in the *IRAS* detector data, and if not perfectly calibrated, would try to fit the gain differences in the detectors by a striped image. The striping builds up in amplitude and sharpness along with the HIRES iterations, as the algorithm refines the "resolution" of the stripes (see Fig. 3.2(a) and (b)).

The IPAC program LAUNDR (Fowler and Melnyk 1990) invokes several one dimensional flat fielding and deglitching techniques. The basic algorithm applied is clamping the background of different scan lines (taken as a low percentile in detector flux histogram for each scan line) to a common level. For the purpose of destriping, the one dimensional algorithm works well for regions with a well-defined baseline, but the result is not satisfactory for regions where structure exists in all spatial frequencies.

3.2.1. *Destriping with Uniform Gain Compensation*

Our approach combines the image reconstruction and the destriping process. Since the striping gets amplified through the iterations, the idea of applying constraints to the correction factors is natural.

Assume footprints in the same leg L suffer from the same unknown gain offset G_L, then

$$D_i^* = G_L D_i \tag{5}$$

is the "true" detector flux, had the detector gain been perfectly calibrated. The G_L's can be seen as extra parameters to be estimated, besides the image pixels f_j. Under a Poisson framework, the maximum likelihood estimate for G_L is

$$\prod_{i \text{ in leg } L} \left(\frac{G_L D_i}{F_i}\right)^{D_i} = \prod_{i \text{ in leg } L} (C_i^*)^{D_i} = 1. \tag{6}$$

in which C_i^* is the gain compensated correction factor. C_i^* is then used in place of C_i in

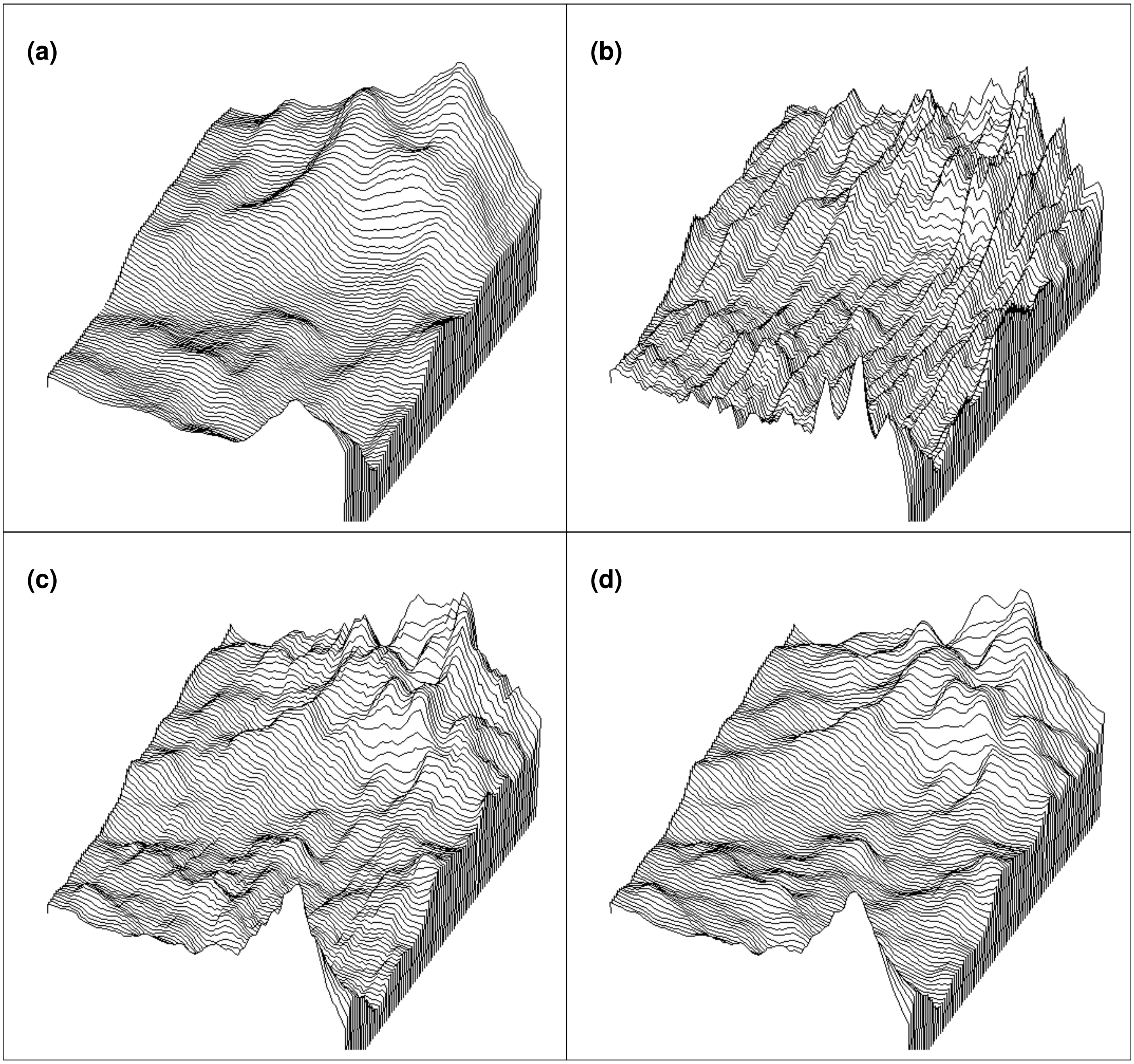

Figure 3: (a). 1st iteration image for a field in ρ Ophiuchus(100 μm band); (b). 20th iteration, standard HIRES; (c). 20th iteration, with uniform gain compensation; (d). 20th iteration, with local gain compensation. Size of image is $1^\circ \times 1^\circ$. Height of surface represents flux. Local gain compensation method produces high-resolution images that are free of stripes, the most common artifact in standard HIRES processing.

Equation (3) to compute the pixel correction factors. A new set of G_L is estimated for every MCM iteration.

This choice of the unknown gain offset G_L minimizes the mutual information between the sets D_i^* and F_i in the leg, i.e. the resulting correction factors C_i^* will extract the minimum amount of information from the stream D_i^*. From the viewpoint of the maximum entropy principle, this is the most reasonable choice.

From another point of view, this strategy works because the procedure of averaging C_i's to get c_j has a smoothing effect on the image, so that the image f_j and estimated flux F_i do not contain as much striping power as the footprints D_i.

3.2.2. *Destriping with Local Gain Compensation*

A further complication lies in the fact that the assumption of a uniform gain offset in a certain leg is only approximately true. Various hysteresis effects (e.g. see Chapter IV of *IRAS* Catalogs and Atlases: Explanatory Supplement (1988)) cause the gain to drift slightly within the 1° range. The more aggressive form of the destriping algorithm estimates the gain offset locally as the weighted geometric mean of the correction factors for nearby footprints, so the estimated gain correction for each footprint varies slowly along the leg. The local gain offset is compared to the global one estimated from the entire leg, and if they differ by more than 10 % then the global value is used, since the gain is not expected to drift that much over a 1° scale, and the variation in computed offset average is most likely due to real local structure. We used an averaging length of $10'$ to estimate the local offset. Because it is larger than the spatial resolution of the first iteration image ($5'$), it is safe to refer the average correction factor on that scale as due to gain offset. The $10'$ length scale is also small enough to capture the drifting behavior of the gain, as shown by visual inspection of both the output images and their Fourier power spectra. Unlike the standard HIRES algorithm (in which stripes are amplified throughout the iterations), the local gain compensation decreases the striping power monotonically to a negligible level after roughly 10 iterations.

One aspect of the local gain compensation method is that the computed correction factors can cause the flux scale to drift slightly. This is solved by requiring an occasional iteration using the standard MCM algorithm to enforce flux conservation. In practice a standard MCM iteration performed at 10 and 20 iterations produced no noticeable re-introduction of stripes.

3.2.3. *Results of the Destriping Algorithm*

Fig. 3 demonstrates the striking effect of the destriping algorithm. Fig. 3(a) shows the first iteration image for a $1^\circ \times 1^\circ$ field in ρ Ophiuchus, which is smooth (blurry). Fig. 3(b) is the 20th iteration image of the field obtained with the standard HIRES algorithm, and is contaminated with strong striping artifacts. A tremendous improvement is seen in Fig. 3(c) which is produced with uniform gain compensation, although some weak stripes are still visible. Finally, using the local gain compensation method gives a stripe-free image, Fig. 3(d). It is also apparent that Fig. 3(d) contains many high spatial frequency features that are absent in 3(a).

3.3. *Subtraction of Zodiacal Emission*

Zodiacal dust emission is a prominent source of diffuse emission in the *IRAS* survey, especially in the 12 and 25 μm bands. The zodiacal contribution to the observed surface brightness depends on the amount of interplanetary dust along the particular line-of-sight, an amount which varies with the Earth's position within the dust cloud. Consequently, the sky brightness of a particular location on the sky, as observed by *IRAS*, changes with time as the Earth moves along its orbit around the Sun. The different zodiacal emission level in different scanlines, if not subtracted, can cause step discontinuities in the images if adjacent patches of sky were observed at different times. Cross-scan destriping helps bring together the background level of scanlines passing through the same local field at different times, but for large scale astronomical studies it is essential to have the zodiacal emission removed.

A physical model of the zodiacal foreground emission based on the radiative properties and spatial distribution of the zodiacal dust was developed by Good (1994). The *IRAS* Sky Survey Atlas (ISSA) (Wheelock et al. 1994) made use of this model and subtracted the predicted zodiacal emission from the detector data before co-adding them. The resulting ISSA images show a strong Galactic background at 60 and 100 μm that is associated with molecular and HI clouds in the Galaxy (e.g. Weiland et al. (1986); Terebey and Fich (1986); Boulanger and Perault (1988); Sodroski et al. (1989); Scoville and Good (1989)). It is desirable for the high-resolution *IRAS* Galaxy Atlas to be sensitive to this large-scale component of Galactic emission.

The zodiacal subtraction is a fairly complicated process, and since HIRES has typically been used for small fields only, the zodiacal emission has been historically treated as part of the local background that is subtracted (and thrown away) during the the LAUNDR step. Fortunately the destriping algorithm offers a way to make use of the zodiacal subtraction effort that went into the ISSA images – we can estimate the zodiacal emission, by comparing input detector data and simulated data from the ISSA images, and calibrate the input data to obtain a background level that is consistent with the ISSA images. The zodiacal emission is taken as a median difference of flux between real and simulated data, computed over a one degree range (the characteristic scale of zodiacal foreground variation), and is then subtracted from the real data. The method therefore only affects the low spatial frequency component of the data, and does not disturb the small scale signals.

When the input ISSA image contains significant striping (at width around $7'$, that is, the distance of neighboring scan tracks, a much larger scale than the HIRES stripes), it is necessary to first smooth the ISSA image with a large kernel ($15'$) before doing the zodiacal subtraction. Otherwise the calibrated detector data would retain the large distance scale offsets, and the gain compensation destriping described in Section 3.2 would not be able to estimate the gain variations correctly and would leave the wide stripes at different flux levels.

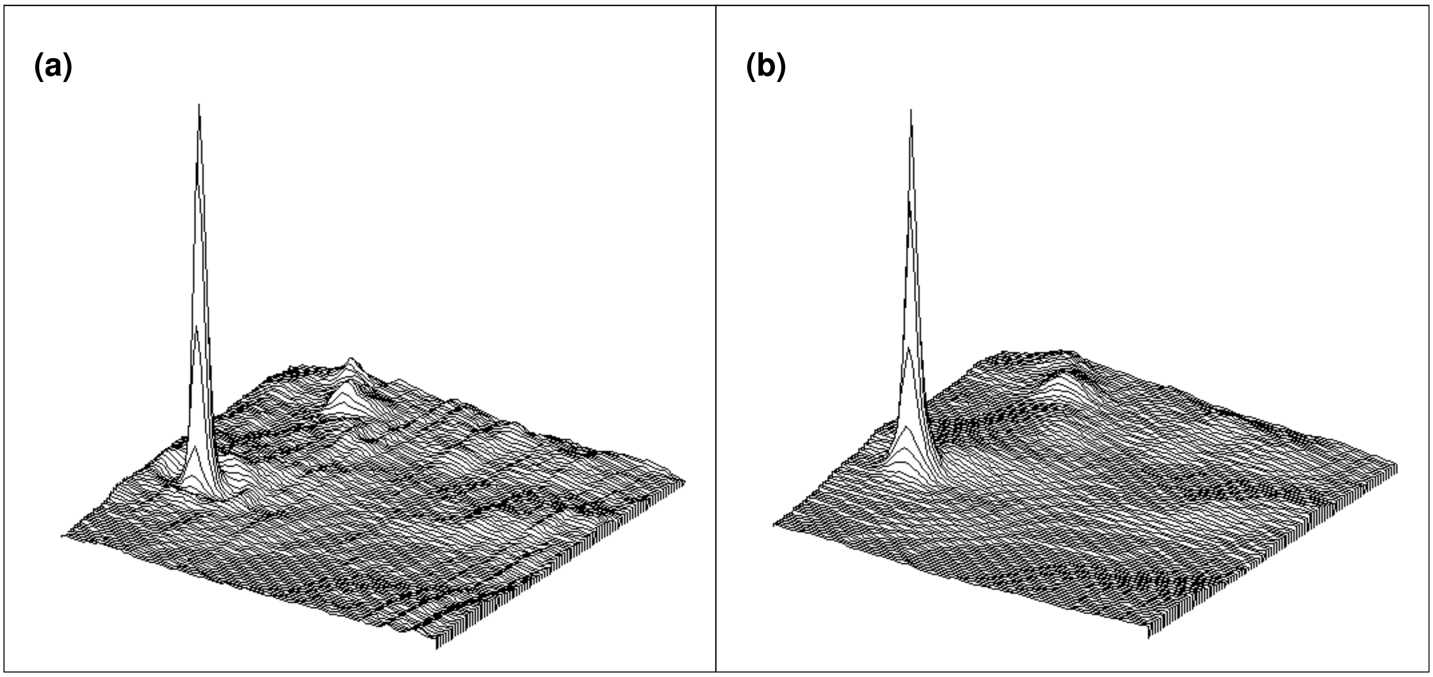

Figure 4: (a). Point source I16293-2422 in ρ Ophiuchus, no ringing suppression; (b). Same field, using entropy prior for ringing suppression. Size of image is $1^\circ \times 1^\circ$. Peak flux in (a) is 3749 MJy/ster, and 3329 MJy/ster in (b).

For validation of this procedure see comparison of surface brightness (output HIRES vs. ISSA) in Section 4.4.

3.4. *De-ringing Algorithm*

For many image reconstruction algorithms, ringing artifacts (or "ripples") appear when a bright point source exists over a non-zero background. The mechanism of the artifact can be understood as the Gibbs phenomenon (a sharp cutoff in high spatial frequency signal incurs ripples in the position domain). Numerous approaches have been taken to reduce this kind of artifact, such as that of regularizing operator in the linear restoration regime (e.g. Zervakis and Venetsanopoulos (1992)), and maximum entropy methods (Press et al. 1992).

A variant of the Log-Entropy MART [3] (De Pierro 1991)

$$f_j^{(k)} = f_j^{(k-1)} + (f_j^{(k-1)})^2 \sum_i \frac{r_{ij}}{F_i^2}(D_i - F_i) \tag{7}$$

was tested on *IRAS* data.

The $(f_j^{(k-1)})^2$ factor in the correction term indicates a Burg entropy metric in the image space: the entropy loss (Burg) involved in changing one's knowledge from f_j to $f_j + \Delta f_j$ is

$$\begin{aligned} -\Delta S &= \sum_j \frac{f_j + \Delta f_j}{f_j} - 1 - \log \frac{f_j + \Delta f_j}{f_j} \\ &= \sum_j \frac{1}{f_j^2}(\Delta f_j)^2 \end{aligned} \tag{8}$$

for small Δf_j, so if we define the distance as $\sqrt{|\Delta S|}$, the metric tensor becomes

$$g_{jj'} = \begin{cases} 1/f_j^2, & \text{if } j = j' \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

The f_j^2 factor then acts to change the covariant gradient vector to contravariant. (See Skilling (1986) for a similar analysis for the Shannon entropy.)

The Burg entropy metric effectively boosts the correction factor for brighter pixels, so the bright point source is fitted better in the earlier iterations, which circumvents the corruption of background propagated from the misfit near the point source.

The prior knowledge signified by using maximum Burg entropy estimation rule has been discussed in Jaynes (1986) and Frieden (1985). According to Frieden (1985), the class of optical objects described by the Burg entropy prior would tend to consist of a relatively small number of randomly placed bright cells, the rest being dim, befitting the bright point source scene we are concerned with.

Suppression of ringing may potentially lead to better photometry determination of the point source through better background determination, and helps solve source confusion problems, which are especially prominent in the Galactic plane.

[3]Multiplicative Algebraic Reconstruction Algorithm

Although the above algorithm gave satisfactory result for some test fields (e.g. see Fig. 4), it suffers from several problems. First, boosting the correction factors for brighter pixels biases the total flux towards higher value, and when this is combined with the destriping algorithm, which essentially is a self-calibrating scheme, gives rise to bootstrapping and uncontrolled growth of flux in the image. This problem can be solved by performing a standard MCM iteration with no destriping and de-ringing applied, before writing out the image. Richardson-Lucy's good property of local flux conservation thus brings back the image flux to the correct level.

Another more serious problem lies more deeply at the heart of the Burg iterative scheme. Since the correction for fainter pixels is damped near bright ones, the Burg iteration is slower at trimming the lobes of point sources. In addition, convergence of faint source near a bright one is also suppressed, along with the formation of the ring.

For these reasons further research is needed to understand the behavior of the ringing suppression algorithm, before it can be incorporated into the production algorithm. In addition, we are investigating an adaptive data splitting method which aims to separate background and point source fluxes during the reconstruction. The Image Space Reconstruction Algorithm (ISRA, e.g. De Pierro (1991)) is also being studied for its potential in ringing suppression.

4. OUTPUT VALIDATION

This section discusses the verification of output image properties.

To test the authenticity of high resolution features produced by the MCM algorithm, Aumann, Fowler, and Melnyk (1990) compared the 60 μm HIRES image of M101 with the *IRAS* Point Source Catalog and previously known H II regions (based on observations at ultraviolet, infrared, and radio wavelengths). Also Rice (1993) examined the structural reliability of HIRES maps for three test galaxies: M51, M33, and NGC 6822 using the following truth tables: (1) a far-infrared KAO map of M51, (2) optical light photographic images of the three test galaxies, (3) four additional types of "high-resolution" maps constructed from independent *IRAS* data, and (4) a simulated map of the radio emission of bright H II regions in M33 constructed from a catalog of 20 cm radio continuum sources in the galaxy.

We've compared the gain compensation destriped images with the original HIRES images for the above fields (and numerous others), and have found good agreement in the reconstructed features (except for the lack of striping). The following sections deal with validation of gain offset recovery, source photometry, source positions, and surface brightness.

4.1. *Validation of Gain Offset Recovery*

To verify the gain offset estimation, a test trying to recover artificially introduced offsets was carried out. A stripe free image was used to generate a set of simulated detector data, and Gaussian-generated gain offsets were applied to the legs. This set of data was fed to the uniform gain offset compensation program, and a scatter plot of the recovered vs. introduced offsets is shown in Fig. 5.

The introduced offsets are Gaussian with standard deviation 0.12. The standard deviation of residual offset after compensating for estimated offset is 0.024, indicating a factor of 5 reduction in striping amplitude (25 in power).

The reconstructed image is stripe free and visually indistinguishable from the input image. This suggests that the uniform gain compensation is capable of its designed goals, and that the residual striping seen in the real data (Fig. 3c) is in fact due to small gain variations within the legs, lending support to the local gain compensation method.

4.2. *Validation of Source Photometry*

To verify the photometric integrity of HIRES images using gain compensation destriping, detector data processed with LAUNDR for the sources M51, M101, and β Pictoris were fed

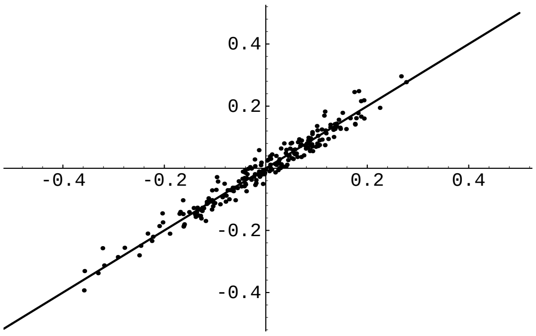

Figure 5: Recovery of Artificially Introduced Offsets. vertical: log of recovered gain offset; horizontal: log of introduced gain offset.

to the gain compensation algorithm and source fluxes at the 20th iteration were compared with results from the standard MCM algorithm. The source fluxes were determined using aperture photometry, by calculating the background level as the median of pixel fluxes in an annulus around the source, and subtracting the background from the total flux within the circle. The maximum percentage difference between the two sets of results is 5.4% (see Table 2 for the comparison).

4.3. *Validation of Source Positions*

To validate the reprojection code and source positions in the output HIRES images, 39 sources near the Galactic plane were checked against the *IRAS* Point Source Catalog. A total of 39 sources between Galactic latitude $-1.7°$ and $1.7°$ ($\times$ two bands) were tested in these longitude intervals: 119° - 129°, 215° - 223°, and 355° - 2°, with 21, 8, and 10 sources in each interval respectively. All sources had fluxes > 1 Jy at 60 and 100 μm. Sources lying within 10 arcmin of each other were excluded. Also some sources close to the Galactic center were excluded, due to a large gradient in the background intensity. For each chosen source, a circular area with radius $5'$ was defined (centered at the PSC position), and the area's flux weighted centroid was taken as the HIRES point source position and compared against the PSC position. For 60 μm band, the distances between HIRES position and PSC position have an average of $9.3''$ and standard deviation $4.9''$, and for 100 μm, $8.3'' \pm 4.4''$.

TABLE 2

Comparison of Source Photometry (Destriping vs. Non-Destriping)

source	wavelength (μm)	preprocessing	destripe?	1st (Jy)	20th (Jy)	diff. at 20th iter. (%)
M51	60	br	no	127.30	131.24	-
M51	60	br	yes	125.53	126.63	-3.5
M51	60	do	no	127.23	130.98	-
M51	60	do	yes	127.23	130.25	-0.6
M51	100	br	no	286.64	303.80	-
M51	100	br	yes	283.28	295.11	-2.9
M51	100	do	no	283.86	299.79	-
M51	100	do	yes	283.86	296.05	-1.2
M101	60	br	no	85.30	85.34	-
M101	60	br	yes	81.72	81.76	-4.2
M101	60	do	no	83.05	84.41	-
M101	60	do	yes	83.08	84.43	$+0.0$
M101	100	br	no	210.95	217.56	-
M101	100	br	yes	207.92	209.97	-3.5
M101	100	do	no	210.70	216.71	-
M101	100	do	yes	210.71	214.37	-1.1
β Pictoris	60	br	no	22.03	21.84	-
β Pictoris	60	br	yes	22.10	20.96	-4.0
β Pictoris	60	do	no	21.58	20.96	-
β Pictoris	60	do	yes	21.51	21.38	$+2.0$
β Pictoris	100	br	no	10.19	10.95	-
β Pictoris	100	br	yes	11.96	10.86	-0.8
β Pictoris	100	do	no	10.45	9.90	-
β Pictoris	100	do	yes	10.42	10.43	$+5.4$

Note: br = baseline removal; do = destripe only, referring to the cross-scan offset done in LAUNDR (not to be confused with gain compensation destriping).

TABLE 3
Comparison of Surface Brightness

position	wavelength (μm)	1st vs. ISSA	20th vs. ISSA
g000.5, +0.5	60	0.024 ± 0.064	0.013 ± 0.100
g121.5, +0.5	60	0.017 ± 0.035	0.008 ± 0.112
g126.5, −0.5	60	0.007 ± 0.040	−0.010 ± 0.119
g218.5, −0.5	60	0.006 ± 0.033	−0.002 ± 0.080
g000.5, +0.5	100	0.031 ± 0.057	0.020 ± 0.083
g121.5, +0.5	100	0.003 ± 0.016	−0.003 ± 0.068
g126.5, −0.5	100	0.008 ± 0.028	−0.004 ± 0.081
g218.5, −0.5	100	0.005 ± 0.016	0.000 ± 0.047

Comparison was done for 1° radius circles centered at positions shown in first column. Differences are represented by mean ± standard deviation of the quantity log(HIRES/ISSA).

4.4. *Validation of Surface Brightness*

To test the surface brightness of zodiacal-subtracted HIRES images, they were rebinned to ISSA geometry (using boxcar averaging) and compared pixel-by-pixel against the ISSA images. The standard deviation of the pixel-by-pixel ($1.5'$) difference is less than 6% for 1st iteration HIRES vs. ISSA, and less than 12% for 20th iteration HIRES vs. ISSA (Table 3). The difference is larger at 20th iteration as the rebinned HIRES images are still sharper than ISSA, while the 1st iteration HIRES images have a resolution similar to the $4'$ to $5'$ of ISSA.

Typical scatter plots of $\log \frac{\mathrm{HIRES}}{\mathrm{ISSA}}$ vs. ISSA intensities are shown in Fig. 6.

5. IMPLEMENTATION

This section gives a detailed account of the technical aspects of a pipeline consisting of coordinated processing on workstations and parallel supercomputers, which produces HIRES images in mass quantities.

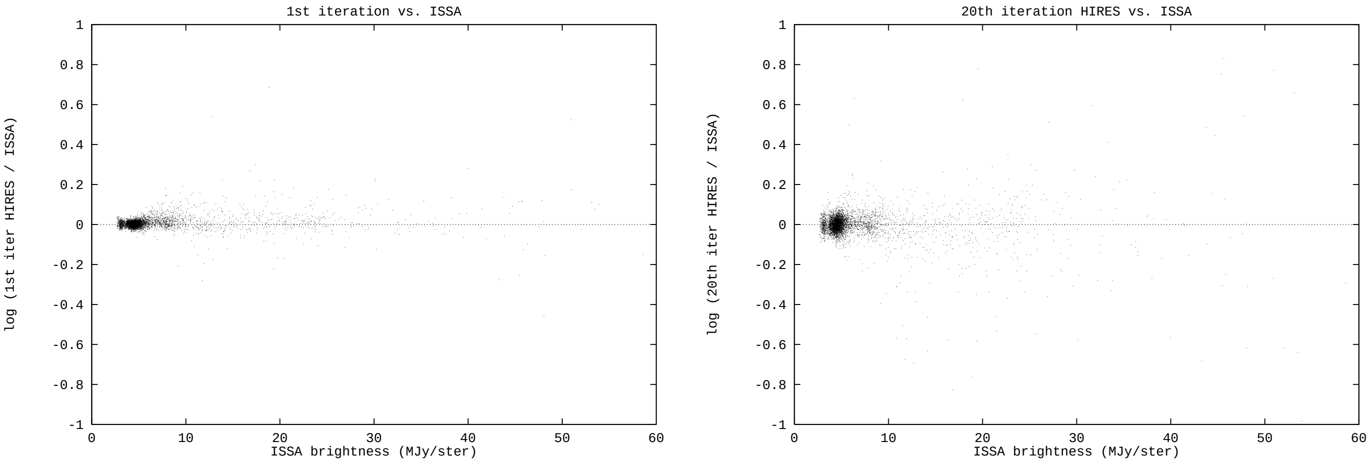

Figure 6: Comparison of HIRES and ISSA Surface Brightness. 1° radius circular area centered at g218.5, -0.5 (60 μm) were compared. Left: comparison of 1st iteration HIRES vs. ISSA; right: comparison of 20th iteration HIRES vs. ISSA. Vertical: log of HIRES / ISSA ; horizontal: ISSA intensity in MJy/ster.

5.1. *Overview of the Production Pipeline*

IRAS detector data, known as CRDD (Calibrated, Reconstructed Detector Data), grouped in $7^\circ \times 7^\circ$ plates, reside in the *"Level 1 Archive"*. The first step in the pipeline for mass production of HIRES images is to extract data covering a specific field with *"SnipScan"* and feed them into *"LAUNDR"* for calibration and various other preprocessing.

We take the $7^\circ \times 7^\circ$ LAUNDRed plates and use the algorithm described in Section 3.3 to subtract the zodiacal background emission. This step requires the corresponding ISSA image as supplement input (SmLAUN in Fig. 7).

Following the calibration and zodiacal subtraction, the detector files are broken into $1.4^\circ \times 1.4^\circ$ fields, and reprojected into Galactic coordinates (from equatorial) if required, with field centers separated by 1 degree (BrkDet in Fig. 7). The factor-of-two overlap is a conservative insurance against discontinuity across field boundaries, as local destriping and different flux bias level will be applied to each small field. $1.4^\circ \times 1.4^\circ$ is also the maximal field

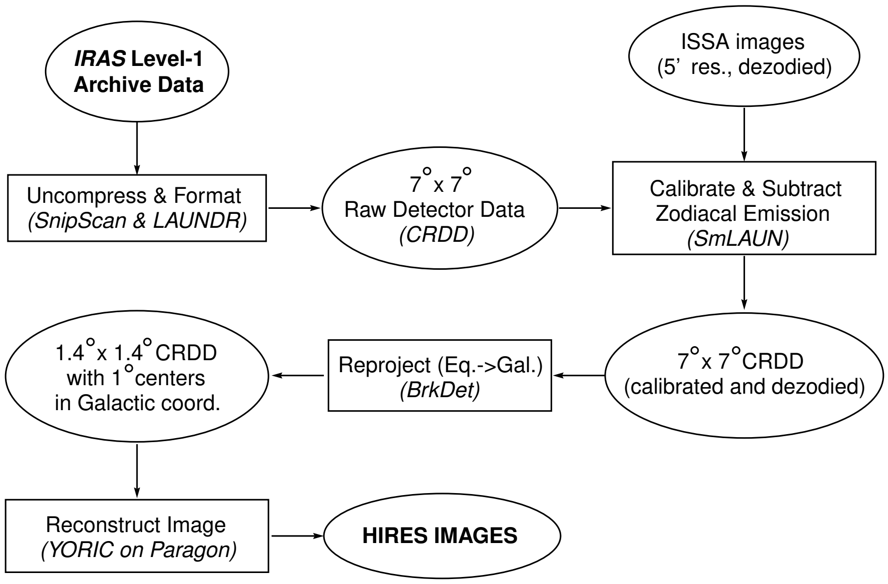

Figure 7: Outline of the HIRES Production Pipeline

size with complete coverage allowed within one Level 1 plate, given the 2 degree redundancy of the plates and arbitrary location and orientation of the small field relative to the Level 1 plate.

All operations described above are carried out on workstations, and the total processing time for one $7^\circ \times 7^\circ$ plate, one wavelength band, averages to 80 minutes on a Sun SPARCstation 10, most of which is spent on decompressing and extracting the data from the *IRAS* Level 1 Archive.

The small field $(1.4^\circ \times 1.4^\circ)$ detector files are then processed into HIRES images, which is done on the Intel Paragon supercomputer. The CPU time taken for this stage is about 100 node-hours [4] for one wavelength band and one $7^\circ \times 7^\circ$ plate. The next section gives a detailed discussion for the parallelization of the Maximum Correlation Method.

The output images are stored to UniTree, a high capacity storage system on the Paragon using HIPPI interface. About 200 megabytes of output data (images and auxiliary maps)

[4] Number of node-hours = number of computing nodes $\times$ number of hours of real processing time.

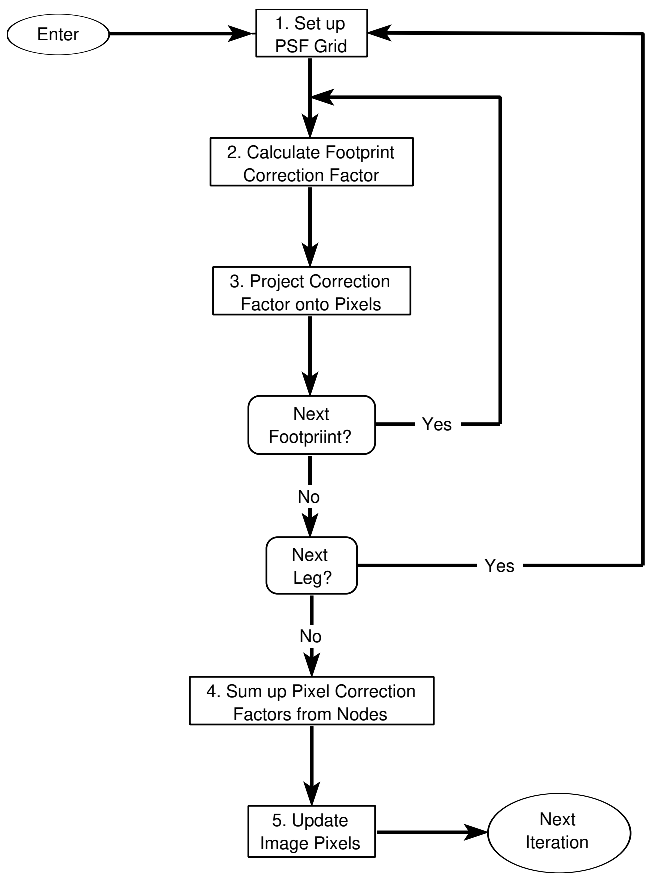

Figure 8: Flow Chart of One Iteration in the Parallel Program

are generated for one band-plate.

5.2. *Parallelization*

A flow chart of one iteration of the parallelized program is shown in Fig. 8.

Profiling [5] a typical HIRES process showed that more than 95 % of the total execution time was spent within the code which calculates the footprint and image correction factors (see Fig. 8). In the parallel decomposition of the problem, each processor takes care of footprints from a set of scanlines. The reasons for doing this are:

[5] Profiling stands for timing analysis of subroutines in the program.

1. Small programming effort. The essence of the original HIRES architecture is left untouched.

2. Footprints in one leg share the same response function grid, except for a translation, which is basically the reason the original code processes the data one leg at a time. Keeping the whole leg in one processor is therefore a natural choice, which minimizes local memory usage.

3. As we discussed in Section 3.2, *IRAS* detectors have gain differences which are especially prominent for the 60 and 100 μm bands. The gain offset can be estimated from correction factors in the same leg, which came from the same detector.

Intermediate disk files for footprint data (D_i) and response function grids (r_{ij}) in the sequential program are replaced by arrays held in memory of the processors (step 1 in Fig. 8), for sake of easier programming and reduction in I/O. This is feasible in the parallel implementation as each processor now holds only a fraction of the entire data set.

Each node calculates the correction factor C_i's for its share of footprints (step 2), and projects them onto the pixels covered by the footprints (step 3). A global sum over all processors for the correction factor c_j's for each image pixel is performed at end of each iteration (step 4), and the weighted average is taken, which is then applied to the image pixel value (step 5).

Decomposition in the image domain was not carried out for the $1^\circ \times 1^\circ$ field, eliminating the need for ghost boundary communication, which would be significant and complicated to code, due to the large size and irregular shape of the detector response function. This helped maintaining the parallel code similar in structure to the sequential one, making simultaneous upgrades relatively easy.

The efficiency of the parallel program depends on the scan coverage of the field processed. The computation time is roughly proportional to the total coverage (i.e. total number of footprints), while the communication overhead is not related to footprints and is only dependent upon the image array size. So the efficiency is higher for a field with higher coverage.

For a large field (e.g. $7° \times 7°$), the detector measurements are broken into $1.4° \times 1.4°$ pieces with 0.4° overlap. Each $1.4° \times 1.4°$ field is loaded on to a subgroup of 8 or 16 processors. The overlap was chosen conservatively so that cropping the overlap after HIRES ensures smoothness at the boundaries. Mosaicked images made from adjacent fields turn out to be seamless to the human eye.

Currently the parallel program runs on the Paragon using Intel's NX communication routines under the OSF/1 operating system. It also runs on the Sandia-UNM Operating System along with the provided communication library (SUNMOS, Maccabe, McCurley, and Riesen (1993)), which is available on Paragon and nCUBE, and provides significant performance increase.

The output images from the parallel computers are compared with those from the standard HIRES program running on a Sun SPARCstation. The differences are well within the range of numerical round-off errors. At the 20th iteration, the standard deviation of (NewImage - OldImage) / OldImage averages to about 10^{-4}.

The global sum operation, which collects pixel correction factors from different nodes, is the primary source of communication overhead in the parallel program.

The executable code was compiled and linked with a math library conformant to the IEEE 754 standard, and the compiler options were fine-tuned to give the best execution speed. For the 60 μm band of M51 (baseline removed data), a time comparison is shown in Table 4.

A speed increase of about 7 times is achieved with 16 processors and 5 times with 8 processors for a $1° \times 1°$ field. Equivalently a 64 square degree field can be processed using 512 nodes, with a speedup factor of 320. For production runs on the Paragon, we customarily use 128 nodes to process 16 small fields simultaneously. Each band-plate would therefore

Table 4

Speed comparisons for 60 μm band of M51

Sun SPARCstation 2	720 sec
Single node of the Paragon	640 sec
8 nodes of the Paragon	137 sec

take roughly 1.5 hours of real time. Various scripts are used to automate the data transfer and program launching.

6. SUMMARY

The parallelization and algorithmic enhancements of the IPAC HIRES program have been described. These efforts have enabled production of HIRES images by IPAC using the Intel Paragon supercomputer.

We are now in the process of producing the *IRAS* Galaxy Atlas, a complete atlas of the Galactic plane ($\pm 5°$ latitude) at 60 and 100 μm with arcminute resolution, as well as maps of the Orion, Ophiuchus, and Taurus-Auriga clouds complexes.

We thank Tom Soifer, Joe Mazzarella, and Jason Surace for their involvement and helpful suggestions during the project. We are grateful to George Aumann, John Fowler and Michael Melnyk for developing the original HIRES program, especially John Fowler who helped with the port of HIRES to the Intel computers by explaining the structure and details of the program, and provided advice throughout the algorithmic developments. Thanks are also due to Ron Beck and Diane Engler who handled numerous HIRES processing requests and are running the Galactic plane image production.

This research was performed in part using the Intel Touchstone Delta and the Intel Paragon operated by Caltech on behalf of the Concurrent Supercomputing Consortium.

REFERENCES

Aumann, H. H., Fowler, J. W., and Melnyk, M. 1990, AJ, 99, 1674.

Beichman, C. A., Myers, P. C., Emerson, J. P., Harris, S., Mathieu, R., Benson, P. J., and Jennings, R. E. 1986, ApJ, 307, 337.

Beichman, C. A. 1987, ARAA, 25, 521.

Bontekoe, T. R., Koper, E., and Kester, D. J. M. 1994, A&A, 284, 1037.

Boulanger, F. B., and Perault, M. 1988, ApJ, 330, 964.

Clemens, D. P., Yun, J. L., and Heyer, M. H. 1991 ApJ Suppl., 75, 877.

Fowler, J. W., and Melnyk, M. 1990, LAUNDR Software Design Specifications, IPAC, Caltech.

Frieden, B. R. 1985, "Estimating Occurrence Laws with Maximum Probability, and the Transition to Entropic Estimators," in *Maximum-Entropy and Bayesian Methods in Inverse Problems*, eds. C. R. Smith and W. T. Grandy, Jr., pp: 133–170, D. Reidel Publishing Company, Dordrecht, Holland.

Good, J. C. 1994, "Zodiacal Dust Cloud Modeling Using *IRAS* Data," in *IRAS Sky Survey Atlas Explanatory Supplement,* Wheelock, S. L., Gautier, T. N. et al., Infrared Processing and Analysis Center, Jet Propulsion Labratory.

IRAS Catalogs and Atlases: Explanatory Supplement 1988, Beichman, C. A., Neugebauer, G., Habing, H. J., Clegg, P. E., and Chester, T. J. (editors) (Washington D.C.: GPO).

IRAS Point Source Catalog, Version 2 1988, Joint IRAS Science Working Group (Washington, DC: GPO).

Jaynes, E. T. 1986, "Monkeys, Kangaroos, and N," in *Maximum Entropy and Bayesian Methods in Applied Statistics*, ed. J. H. Justice, pp: 26–58, Cambridge University Press.

Lucy, L. B., 1974, AJ, 79, 745.

Maccabe, B., McCurley, K. S., and Riesen, R. 1993, SUNMOS for the Intel Paragon, ftp://cs.sandia.gov/pub/sunmos/doc.

De Pierro, A. R. 1991, "Multiplicative Iterative Methods in Computed Tomography," in *Mathematical Methods in Tomography*, eds. G. T. Herman, A. K. Louis and F. Natterer, pp: 167–186, Springer-Verlag.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. 1992, Numerical Recipes in C, 2nd Ed., 823.

Rice, W., 1993, AJ, 105(1), 67.

Richardson, W. H., 1972, J. Opt. Soc. Am., 62, 55.

Scoville, N. Z., and Good, J. C. 1989, ApJ, 339, 149.

Skilling, J. 1986, "The Cambridge Maximum Entropy Algorithm," in *Maximum Entropy and Bayesian Methods in Applied Statistics*, ed. J. H. Justice, pp: 26–58, Cambridge University Press.

Sodroski, T. J., Dwek, E., Hauser, M. G., and Kerr, F. J. 1989, ApJ, 336, 762.

Soifer, B. T., Houck, J. R., and Neugebauer, G. 1987, ARAA, 25, 187.

Snell, R. L., Heyer, M. H., and Schloerb, F. P. 1989 ApJ, 337, 739.

Surace, J. A., Mazzarella, J. M., Soifer, B. T., and Wehrle, A. E. 1993, AJ, 105, 864.

Terebey, S., and Fich, M. 1986, ApJ, 330, L73.

Terebey, S. and Mazzarella, J. (eds.) 1994, Science with High Spatial Resolution Far-Infrared Data, (Pasadena: JPL).

Weiland, J. L., Blitz, L., Dwek, E., Hauser, M. G., Magnani, L., and Richard, L. J. 1986, ApJ, 306, L101.

Wheelock, S. L., Gautier, T. N. et al. 1994, *IRAS* Sky Survey Atlas Explanatory Supplement, (Pasadena: JPL).

Wood, D. O. S., Myers, P. C., and Daugherty, D. A. 1994, ApJS, 95, 457.

Zervakis, M. E., and Venetsanopoulos, A. N. 1992, Multidimensional Systems and Signal Processing, 3(4), 381.