EVALUATION OF IMAGE QUALITY FOR SIM, USING DIFFERENT (U,V)-PLANE COVERINGS AND THE SIMSIM SOFTWARE

B-G Andersson

DRAFT REPORT 980714

Abstract:

A study of the image quality of reconstructed images which can be expected using imaging-mode observations with the SIM instruments has been performed. The STScI software "SIMSIM" was used to study, primarily, the consequences of eliminating short spacings in the (u,v)-plane coverage. The results agree with arguments based on simple spatial-frequency arguments and show that for simple and likely combination of source types within the field-of-view, the exclusion of short interferometer spacings can lead to sever degradation in the determination of source parameters such as size, photometry and (faint) source count. For any permanent exclusion of short spacings there will be sources for which significant degradation in the measurements will result.

EVALUATION OF IMAGE QUALITY FOR SIM, USING DIFFERENT (U,V)-PLANE COVERINGS AND THE SIMSIM SOFTWARE

Evaluation of Image Quality for SIM, using different (u,v)-plane coverings and the SIMSIM software

Abstract

A study of the image quality of reconstructed images which can be expected using imaging-mode observations with the SIM instruments has been performed. The STScI software "SIMSIM" was used to study, primarily, the consequences of eliminating short spacings in the (u,v)-plane coverage. The results agree with arguments based on simple spatial-frequency arguments and show that for simple and likely combination of source types within the field-of-view, the exclusion of short interferometer spacings can lead to sever degradation in the determination of source parameters such as size, photometry and (faint) source count. For any permanent exclusion of short spacings there will be sources for which significant degradation in the measurements will result.

Introduction

In order to study the impact of partially unfilled (u,v)-plane sampling, the STScI SIM simulator software (SIMSIM; Boeker & Allen) has been used to 1) construct model sources and 2) produce simulated observations (including reconstruction) by the SIM instrument. In this study only the available sources in the SIMSIM package were utilized. A number of sources and (u,v)-plane coverages were considered. For each source simulation a single 500nm wide channel was used over a 128x128 pixel grid. The "baseline" simulation had all radial "slots" filled in in the (u,v)-plane (i.e. every 0.5 m from 0.5 to 10.0) and an angular separation of 20˚ (by visual inspection - and as the results will show, below - this configuration produced a very clean "dirty beam"). To evaluate the image degradation with missing short spacings, three simulations were run for each source with a central "hole". These had as their shortest baseline 2.0, 2.5 & 3.0m (for the baseline case of a field of point sources larger steps were employed). For comparison purposes a fifth case was run with all radial points used but for which the angular step size was increased to 60˚. Since the main impetus for the present study was the issue of what short spacings are necessary to produce reliable images, but not whether, or to what extent, the angular step size may be restricted, these cases are more to be thought of as comparisons for the effects of lessening the total number of (u,v)-points than conclusive results in and of themselves.

It was found that the metric of "recovered flux" yielded the best and most convenient measure of image quality in the present study. This measure, of course, presumes knowledge of the "true" flux. The fact that these are available to us sets this study somewhat apart from earlier (and more comprehensive) studies of interferometric image quality such as Cornwell et al. (1993).

Sources

The sources considered were: 1) A field of ten, point like, stars. 2) A 0.1" (FWHM) disk with either a power-law spectrum of spectral index: -0.5, or a 3000K thermal spectrum, and 3) a combination of stars and a disk. Two sub-cases of this last one were run; a) with equal number of counts from the stars and from the disk and b) with a factor 100 less counts from the disk than from the stars. For all cases a fairly large number of total counts were used in order to emphasize the reconstruction part of the problem. Note also that in the figures below, the abscissa is not numerical. In all cases the CLEAN algorithm was used for reconstructing the image from the "measured" visibilities. Within each simulation set (source) a constant number of CLEAN steps was employed, typically 1000. A visual inspection of the residual maps (as presented in SIMSIM) was made to assure that the

residuals had reached the noise level. Secondly, the "difference" parameter in SIMSIM was required to have reached its asymptotic value.
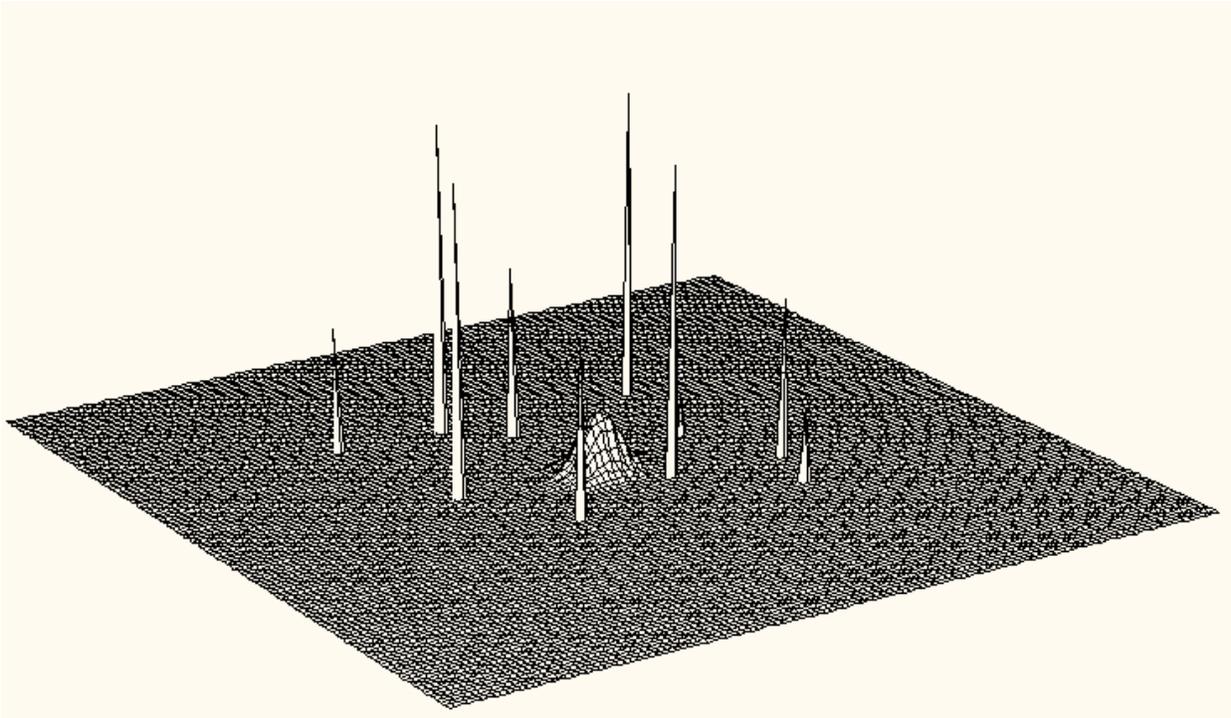
Figure 1. A surface plot of the combination of a strong disk and ten stars in the field (case 4 below) illustrates the kind of sources used in the present study. Note that the brightest "stars" have been truncated at fluxes of 1x10^5 counts in order to show the disk source more clearly. The total number of counts in the disk as well as the sum of all stars are both 1x10^6

Although only a limited number of sources were used, the results already show that for any (u,v)-plane coverage with a significant hole at the center there are source combinations and scientific problems which will be severely handicapped.

Evaluation

The question of how to, quantitatively, evaluate the simulation results is nontrivial. As has been pointed out in the literature (e.g. Cornwell et al, 1993), the best metric is highly dependent on the source and application of the imaging. Therefore, several different methods/metrics were tried. Here, before presenting the simulation results, a few of the possibilities will be briefly discussed. In all cases, discussed herein, we have the somewhat unrealistic advantage of being in possession of the "true" source structure (i.e. the input model). Obviously, in a real observation this is not the case. This fact should be borne in mind.

Image Size:

(u,v)-cover.	Disk FWHM
Model	6.2
all baselines	6.2
b 2.0m	5.6
b 2.5m	5.2
b 3.0m	5.7
all b, $\Delta\theta=60°$	6.1

Table 1 FWHM of reconstructed Disks

For extended sources one possible metric is the extent and shape of the recovered source compared to the input. For the disk simulations this measure was used. In figure 2, I have plotted cuts through the central three rows of pixels from this simulation family. As can be seen, the "fully" sampled (u,v)-plane case diligently reproduces the model source. Even the case where baseline up till 1.5m have been excluded reproduces the central part of the disk. Note however that at the low levels on the outside of the source, this case under estimates the source brightness. For successively larger "central hole" both the source brightness and size becomes increasingly poorly determined. The FWHM widths of the different sources progress as follows (in pixel units):

Because several of the sources used (as well as sources on the sky) are point like and we used post-CLEANing images, the shape of the source is a poor choice of sole image quality metric when the field contains point sources.
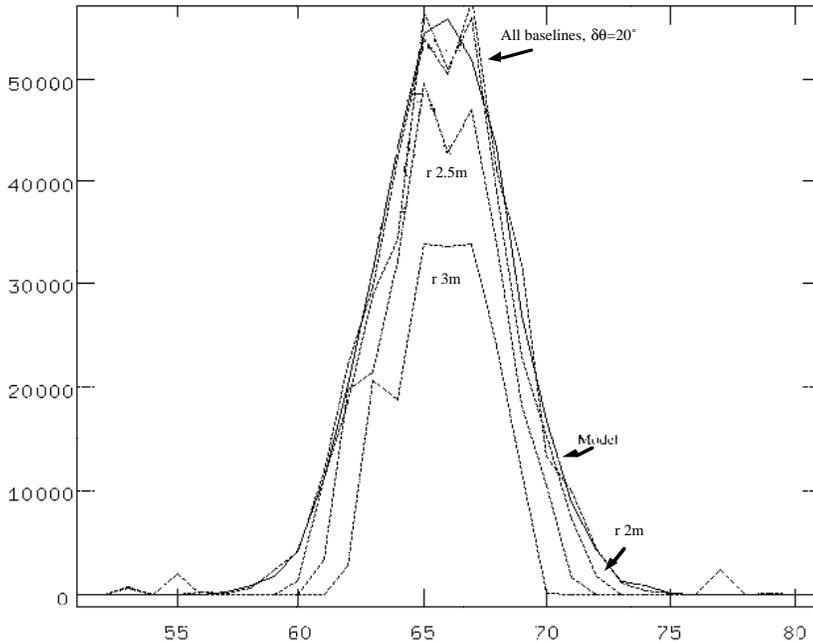
Figure 2. *Cuts through a disk as simulated by SIMSIM for different (u,v)-plane coverages. Full line is for the input model, dashed lines represent the different observation simulations.*

Recovered Flux:

A second possible choice for a metric of quality of image reconstruction is how well the observation reproduces the flux of the source. I have used IRAF, aperture-photometry, routines to evaluate this measure. In crowded fields this method can run into severe problems when a neighboring source falls within the annulus used to evaluate the "sky background" for the source. Software (e.g. DAOPHOT) does exist that will address such sources - i.e. globular clusters - but I have not applied these here. Even for less crowded fields the choice of aperture size is somewhat arbitrary. Usually a set of increasingly bigger apertures are employed and the source flux is taken to be the asymptotical value (accounting for background and neighboring sources) of the fluxes. In order to allow for a somewhat more realistic evaluation of the images (rather than picking the sources I knew to be present) I have used a somewhat "blind" procedure. After having set the search and photometry parameters to reasonable values I allowed the IRAF routines to run and

accepted the sources they found. This list could then be used to both, address the issue of recovered flux and dynamical range.

Specifically, I have employed the IRAF routines DAOFIND to find the "detected" sources in the (SIMSIM) output images and QPHOT to perform the aperture photometry. In QPHOT apertures of 1,2,3,5,10,15 and 20 pixel radius were used. The smaller ones were used for the star like sources while the larger ones were used for the disks.

Figure 3 shows the recovered flux for the case of a field of ten stars with a bright disk in the same field. As can be seen the recovered flux behaves as would be expected from simple spatial frequency arguments.

10 stars (Σcounts=10^6) + strongdisk (Σcounts=10^6)
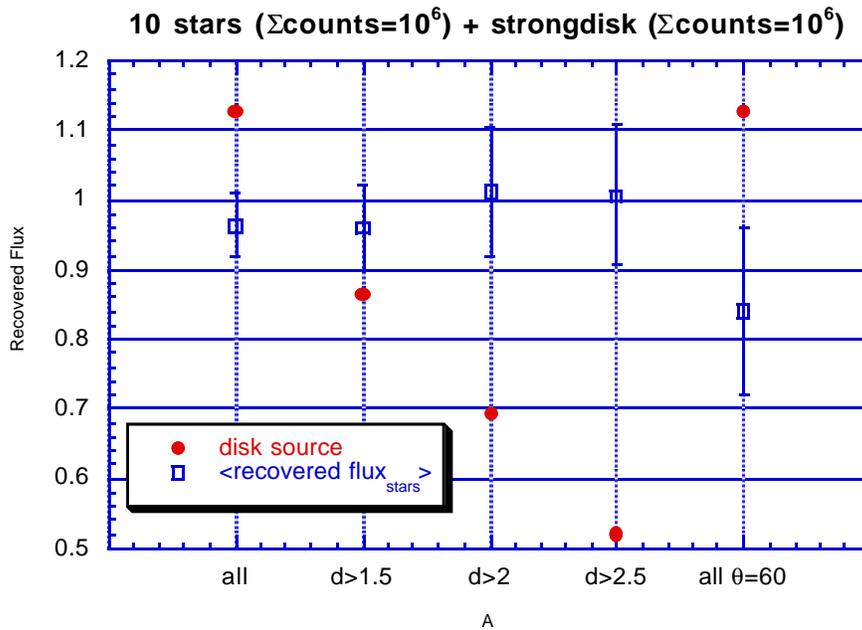
Figure 3. *The recovered flux for a simulation of a field containing ten stars and a "strong" disk (number of counts from disk equal to the number of counts form the stars) For each source in the field I have taken the ratio of the flux measured in the input model and that measured in the reconstructed image. The open squares represent the average of these ratios for the ten stars while the filled circles represents the extended disk source. (see further discussion below)*

Dynamical Range:

The dynamical range is defined by Cornwell et al. (1993) as "the peak brightness in an image to the off-source error level". I will here modify this definition somewhat and use the ratio of the peak brightness in the image to the brightest, non-real (since we use theoretical models as inputs) source detected by the source finding algorithms. In figure 4, I plot the Dynamical range for the "ten star plus weak disk case". Also plotted is the ratio of the detected flux for the weakest source in the original input image to the flux of the brightest spurious source detected. I will refer to this as the detection margin. Both of these measures are obviously somewhat arbitrary numbers in that they depend on the flux distribution of the input sources. However, (figure 3) both measures can provide useful information.
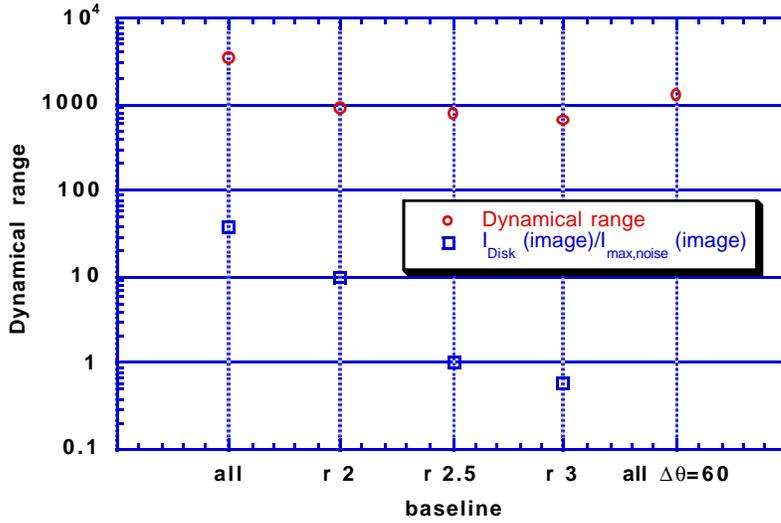
Figure 4. *Dynamical range and ratio of flux detected for the weakest input source to brightest spurious source for the case of a ten star + weak disk input model. Note that for the case of sparse azimutal coverage, the disk source was not unambiguously detected.*

Fidelity:

A number of specific formulae were tried to use the metric of "fidelity" as suggested by Cornwell et al. (1993). The original formulation, suggested by those authors;

$$f_{pix} = \frac{I_{pix}}{\sqrt{\left(T_{pix} - I_{pix}\right)^2}}$$

where T and I are the model (true) and reconstructed images, was found to unsatisfactory since a large number of pixels are blank (in both the model and reconstructed images) and the denominator for these hence, of course, diverges. The most promising formulation for our application seemed to be the simple modification:

$$f_{pix} = \frac{I_{pix}}{1 + \sqrt{\left(T_{pix} - I_{pix}\right)^2}}$$

This metric, when averaged over the images (using IMSTAT in IRAF), did yield results consistent with visual inspections and the metric of "recovered flux". However, for all images a number of outlier pixels occurred (presumably the locations of the stars) which, unless selectively suppressed, tended to bias the results towards higher (and seemingly somewhat random) values. Also, the variance around the mean fidelity was always much larger than the mean and also much larger than the differences of the means between different simulations.

Figure 5 shows the results for one of the sources cases.

Data 6
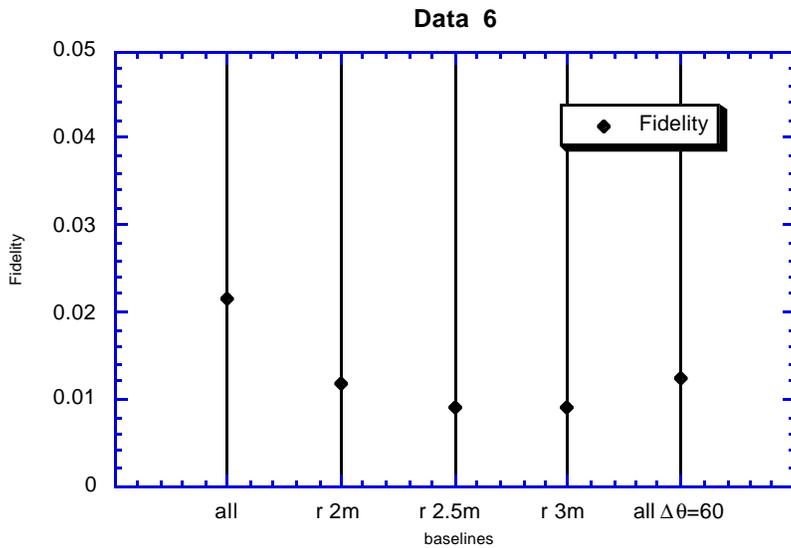
Figure 5. Source fidelity as a function of (u,v)-plane coverage for a combination of ten stars and a "weak" disk. Although the fidelity-averages over the image acts as expected the standard deviations over the image is much bigger than the simulation-to-simulation variations

Results

As indicated above, three main source classes were investigated, collections of point sources, extended sources (Gaussian disks) and a combination of the two. Several sub-classes of each were considered. Here I will discuss the SIM simulations of each class in turn.

Case 1:

As a sanity check, the first case we will consider is that of a field of point sources. The source example which SIMSIM provides is one of 10 sources in the field. Total flux is 10^6 photons. As can be seen from figure 6 the recovered flux is very close to total for most (u,v)-plane coverages. Indeed, this would be expected from "hand waving arguments".
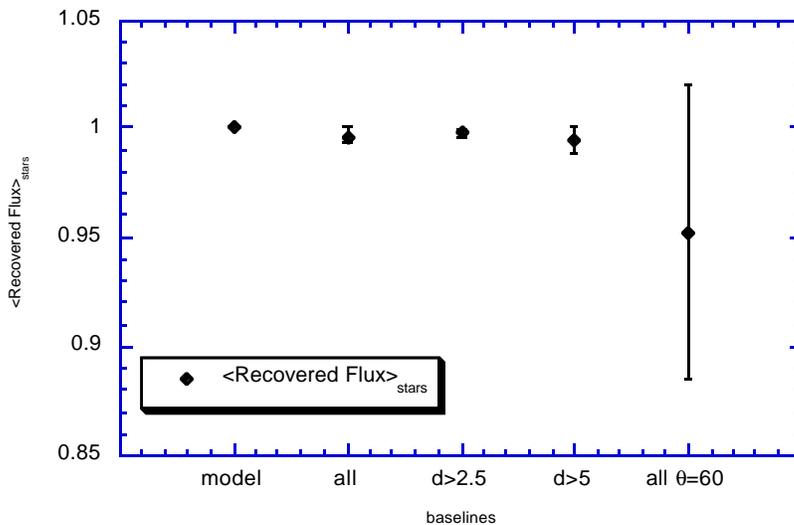
Figure 6. Recovered flux for ten stars in the FoV. The plotted values are the averages of the ratios between model (input) and reconstructed fluxes. The error bars correspond to the star-to-star variations.

Case 2:

A Gaussian disk of 0.1" FWHM of total flux 10^6 photons. Spectral distribution given by a power law of index -0.5. (Note, however that a single 500nm wide channel was used). A successive drop-off in the recovered flux can be seen for increasingly big central "holes". Already when the first baseline is 2m the recovered flux has fallen below 90%. Not that for the case where only baselines between 1.5-2m have been deleted (e.g. simulating a successful combination of SIM with HST imaging) almost all the flux is again recovered. Note also table 1 and figure 2, above, for the effect on the recovered size of the disk as a function of (u,v)-plane coverage.
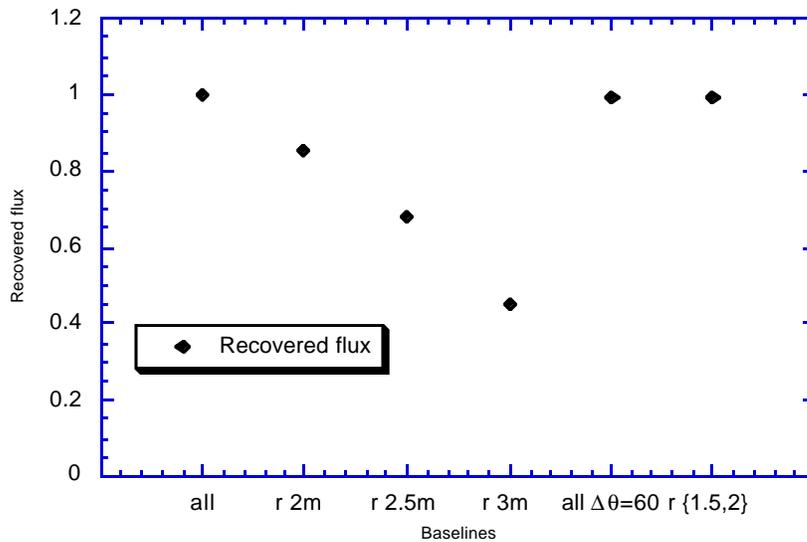
Figure 7. A disk of 0.1" FWHM with a power law spectrum.

Case 3:

A Gaussian disk of 0.1" FWHM of total flux 10^6 photons. Spectral distribution now given by a 3000K blackbody. The results are very similar to the power-law disk, but do show some differences, most probably due to the differences in effective baselines (i.e. baselines measured in units of wavelengths) included.:
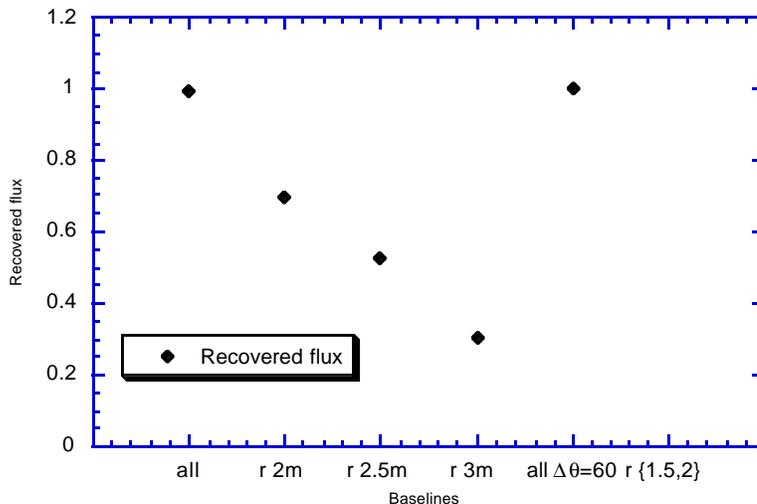
Figure 8. Same disk as in figure 6, except a blackbody spectrum (T=3000K)

Case 4:

A combination of ten stars in the field and a disk (as in case 2, a power law spectrum was used) of equal fluxes (i.e. the total flux in the stars is 10^6 photons and the total flux in the disk is 10^6 photons; see below, however). Both the stars and the disk behaves, broadly, as when each source is alone in the field. The fact that the recovered flux for the stars with a central hole 2.5m exceeds 1.0 can be attributed to grating rings from the disk source which coincide with the location of several of the stars. Source detection/count now becomes an issue and as figure 10 illustrates the results are very sensitive to the (u,v)-plane sampling.
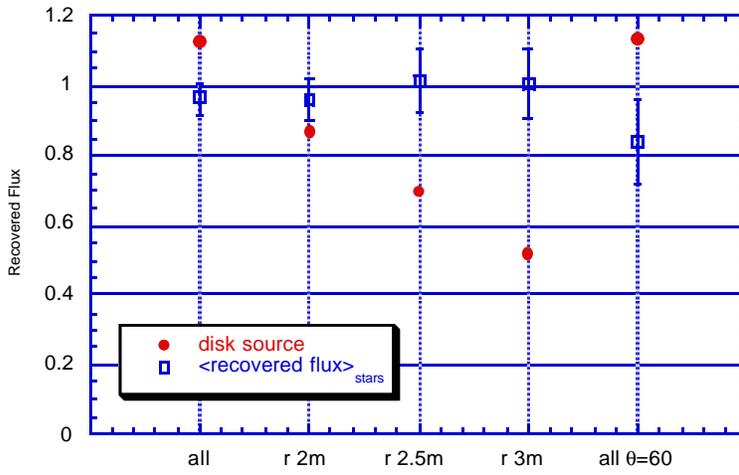
Figure 9. Ten stars and "bright" disk. As above the recovered flux for the stars represents the average of the stars and the error bars the star-to-star variation
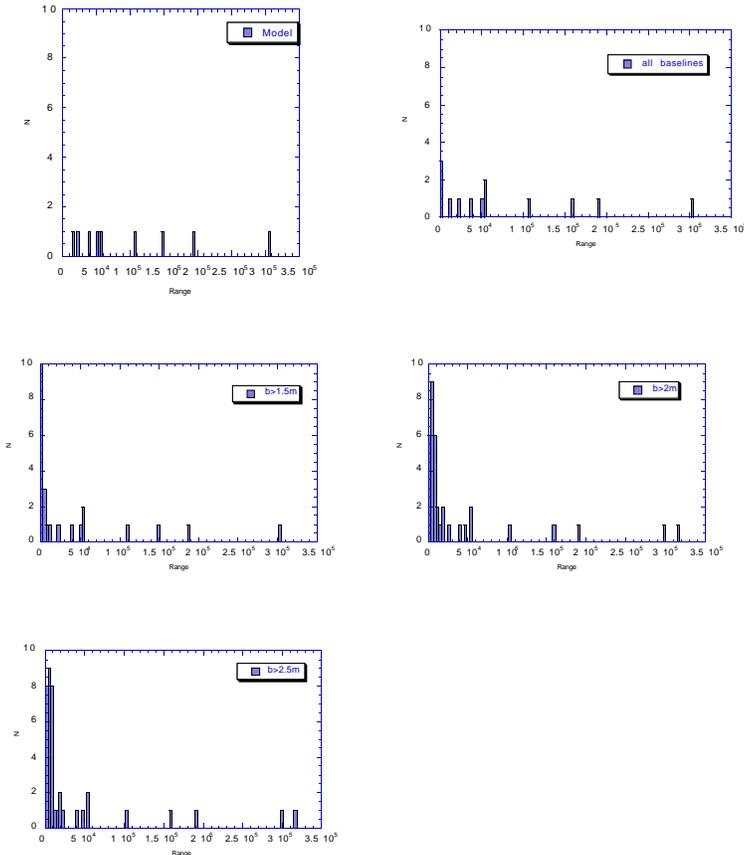
Figure 10. Using the autonomous source finding and photometry routines in the IRAF APPHOT package the reconstructed images in case 4 were analyzed for point sources. The histograms show the statistics of the point source photometry. A region if 10x10 pixels centered on the disk source was excluded from the analysis. As would be expected, the number of weak, spurious sources increases dramatically when the (u,v)-plane coverage is less and less complete. If the purpose of the observation had been to study the source statistics, e.g. in an attempt to establish an IMF in a stellar cluster, the results would suffer if short baselines were missing.

In order to begin investigating the influence of integration time on the source reconstruction the source of case 4 was scaled by a factor 0.1 and 0.001 using IRAF routines and subsequently, again, run through the SIM simulation. In figure 11 the measured half-power widths of the disk are plotted for the different source fluxes and different (u,v)-plane coverages. In figure 11 the behavior of the dynamical range for point sources is illustrated for this sequence of simulations.
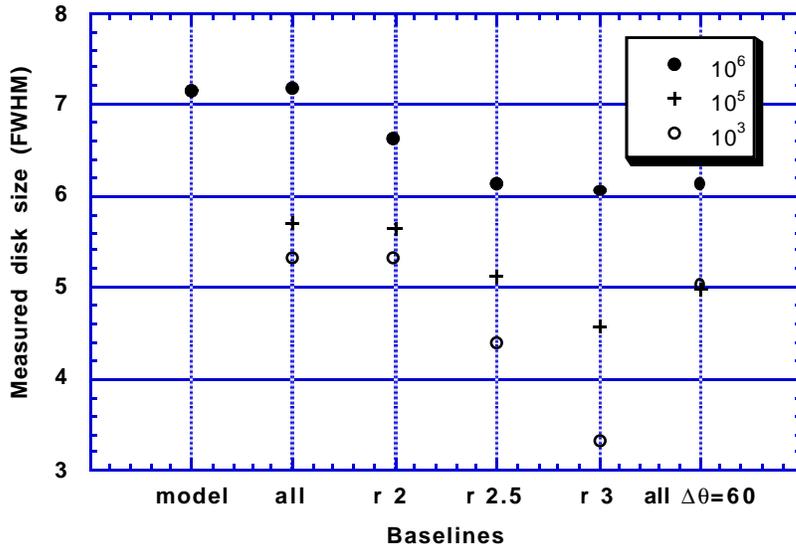
Figure 11. *The measured size of the disk in a combined-source field degrades dramatically, relative to the input model, as the total flux - or equivalently, the integration time - in the image decreases. Here a combination of 10 stars and an extended disk source with 10^6, 10^5 & 10^3 counts each are used to illustrate this effect.*
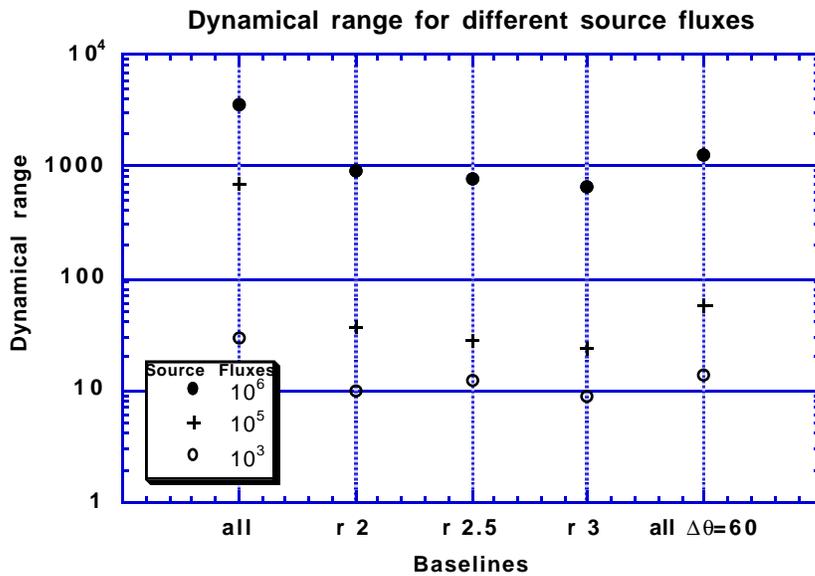
Figure 12. *The dynamical range of the bright disk plus star simulations, defined as the flux of the brightest source in the image divided by the brightest spurious source in the image is shown for different baseline coverages. Again, a 10x10 pixel region surrounding the disk was excluded in this analysis.*

Figure 13 illustrates how the dynamical range depends on the flux level for each simulation type. As can be expected, the dynamical range increases with increasing total flux somewhat slower than linearly, but more rapidly than (photon-noise characteristical) square root dependence. A more detailed analysis than the present is needed to elucidate the origin of the detailed behavior.
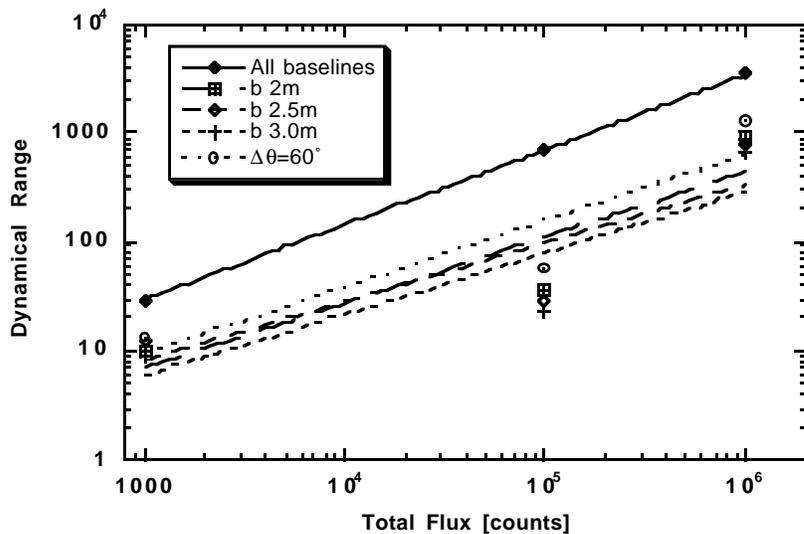
Figure 13. *The dynamical range in the "case 4" simulations decreases rapidly with lowered total flux. If we approximate the dynamical range by a powerlaw we find exponents between 0.53 and 0.69. This is slightly worse than if the dynamical range was dominated by simple photon counting statistics (which would have predicted 0.5)*

Case 5:

A combination of ten stars in the field and a disk (as in case 2) of 100 times smaller flux (i.e. the total flux in the stars is 10^6 photons and the total flux in the disk is 10^4 photons). The disk is now very sensitive to the (u,v)-plane coverage. This is probably due to the fact that the CLEAN algorithm was used in the reconstruction of the image. Clearly, if the stars are bright, the first few clean components will pick up stars and hence any side-lobes on the dirty beam will cause photometric errors in the low level disk distribution. At the present time, SIMSIM does not support other reconstruction algorithms, such as MEM.
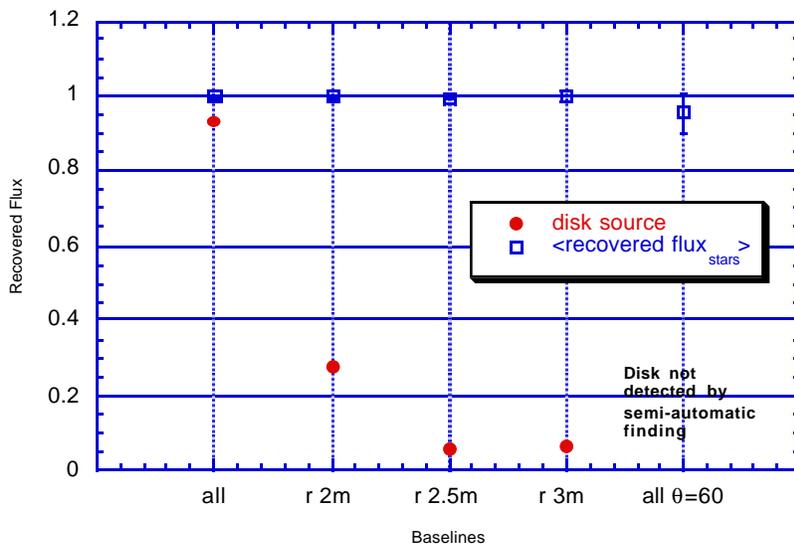
Figure 14. *Ten stars and "faint" disk. As above the recovered flux for the stars represents the average of the stars and the error bars the star-to-star variation.*

For comparison between cases 4 and 5, the dynamical range for point sources in the image (ratio of brightest source flux to brightest spurious source detected) is plotted in figure 15 while figure 15 plots the size of the recovered disks. As might be expected, for a relatively bright disk, poor (u,v)-plane coverage causes the point source detection characteristics to degrade, whereas for a faint disk the determinations of the disk parameters suffer.
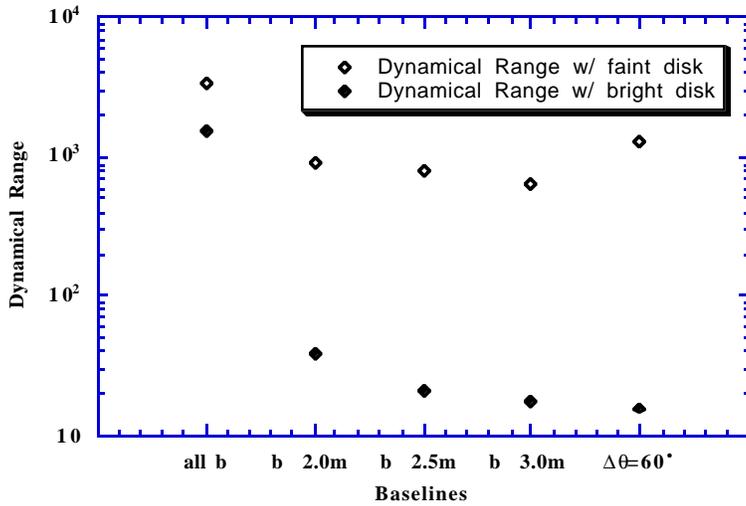
Figure 15. *The dynamical range of the two disk plus star simulations, defined as the flux of the brightest source in the image divided by the brightest spurious source in the image is shown for different baseline coverages. Again, a 10x10 pixel region surrounding the disk was excluded in this analysis.*
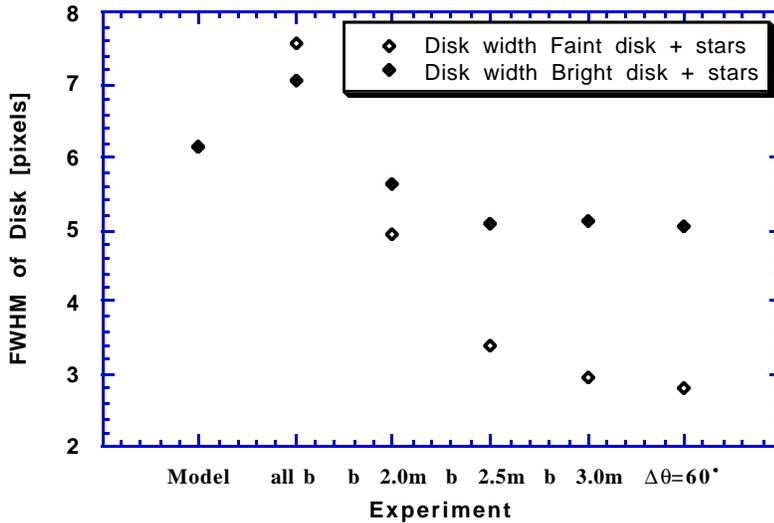
Figure 16. *The measured width of the disk in the simulations (FWHM) is shown for the different baseline coverages. Note that, as would be expected, the error in the disk parameters are worse when the stars in the field dominate the flux as opposed to the result for the dynamical range above (fig. 11)*

Discussion

As can be seen from the above cases, and as would be expected from simple spatial frequency arguments, extended sources suffer losses of their recovered flux when the shortest spacings are missing. However, also point-like sources can be affected by missing (u,v)-coverage through the effects of a complicated "dirty beam", particularly in instances where there are both point like and extended sources present. The same effect would be present in the case of crowded fields, since the overlap of the diffraction patterns for neighboring sources is similar to those for extended sources. Note that even though the source photometry might be improved by the application of crowded field photometry software (e.g. DAOPHOT), the fundamental problem is not alleviated since the CLEANing precedes the photometry. As the two last cases illustrate, there are source combinations of stars and extended sources that will introduce significant errors in the parameter determinations for either source type, when short baselines are missing, even in the case where the total photon count in the image is high. When the disk is relatively weak, such that the first few CLEAN components are all located on the stars, the disk parameters such a size and photometry will primarily suffer. When, on the other hand, the disk is bright enough that CLEAN picks it up early, the dynamical range in the point source observation is severely compromised. As figure 15 shows, already if only the first three baselines {0.5m,1.0m,1.5m}are missing the dynamical range is decreased by about a factor 30 for this case.

In figure 17 I have plotted the approximate size of the central "hole" (i.e. size of the missing short baselines) at which 25% of the flux is lost for the different extended sources considered above. As can be seen there are "realistic" sources for which no loss of (radial) (u,v)-plane coverage can be accepted (bright stars and faint disk within the same FoV). For other extended sources (e.g. isolated disks), the first meter of (u,v)-plane coverage can be relinquished with only "minor" loss of photometric source size accuracy.
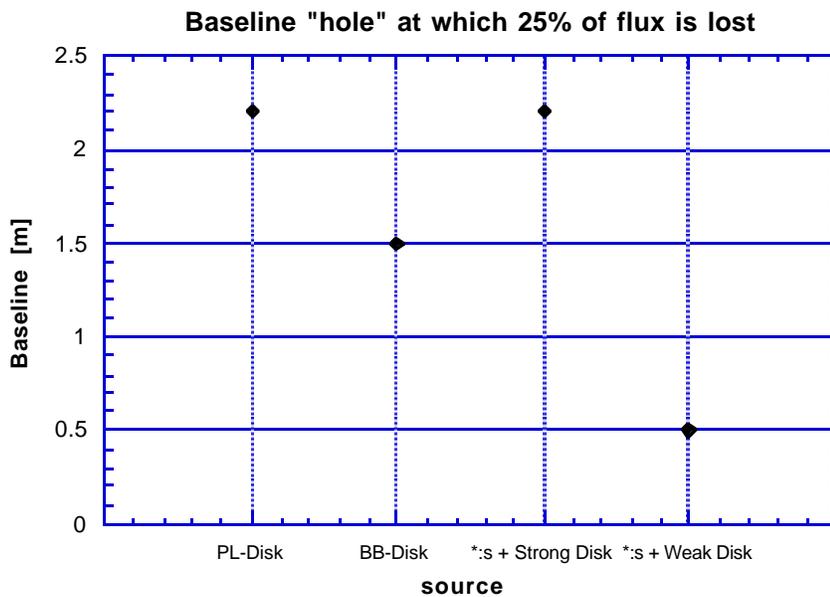
Figure 17. The size of the central (u,v)-plane hole for which 25% of the source flux is lost is plotted for the sources considered in this report.

To what extent single dish (i.e. HST) imaging can be used to fill in the missing shortest spacings in the (u,v)-plane depends on several factors. Most importantly if the spectral pass band of the SIM and single dish observations can be made to match. For broadband/continuum measurements it seems likely (though I have not in this study tried to address this issue) that this should be possible. However, as Ron Allen has pointed out to the SIMSWG, there are many sources of interest to the SIM community whose inherent spectral characteristics (e.g. emission-line sources) makes it very doubtful that a good enough spectral passband match can be made between SIM and HST for the co-addition of visibilities to be feasible.

Further work

Several extensions of this preliminary study should be pursued to quantify to what extent any given loss of short spacings will limit the imaging applications for SIM. A draft list would contain:

- What influence would other deconvolution algorithms have on the above results?

   Could some of the problems encountered in the "combined source" cases above be circumvented by using MEM(not yet implemented in SIMSIM) or more optimized versions of CLEAN?

- Can "crowded-field" photometry improve the source count/dynamical range behavior?

   DAOPHOT and other crowded-field photometry packages treats source photometry somewhat like CLEAN in that a detected source is subtracted before the next one is measured. Would that improve the photometry?

- How would the results differ if the size of the extended source changes?

   If the disk size becomes smaller the number of CLEAN components dedicated to it should decrease, will that help in terms of dynamical range?

- Given a specific SIM architecture, and preferably more realistic and *specific* sources, what limitations can be set on the determinations of derived source parameters?

- Some work on extensions and modifications to SIMSIM will need to be pursued. These upgrades would include:

   MEM image reconstruction

   A more flexible source construction (e.g. total flux)

   Bug fixes.

Conclusion

The main issue which I have sought to address herein is the vague question of "what is the longest, shortest spacing which can be accepted from a scientific point of view for SIM?" Clearly that will depend on what amount of image/photometric degradation can be accepted as well as to what extent *a priori* source selection can be accepted. The present study is not comprehensive enough to address, quantitatively, what those restrictions will be.

However, if the goal is that *any* source should be reliably observable if the required observing time can be allocated then the answer is that **no** significant baseline range can be permanently unattainable.

Hence, if no source type is to be *a priori* excluded from quantitative observations, most (if not all) possible short baselines need to be observable. Conversely, if a range of short baselines is permanently excluded from observations, limitations will be imposed on the sources that can be observed and/or the quantitative determination of source parameters. This conclusion is probably somewhat stronger than stated here, since the detailed structure for sources observable with SIM is unlikely to be simple even if a pre-selection is made. The SIMSIM software, given some upgrades and new sources can be used to address these limitations. Most profitably, a given, restricted, SIM architecture should be identified for which these limitations could then be investigated.