# Chapter 2

# Mock Observations

We use results from mock observations to determine the strategies for studying extragalactic sources with SMA and ALMA. We can have realistic estimates and understand their limitations, which are not available from theoretical calculations. *Miriad*, a radio astronomy package, is the main code to carry out our mock observations and analyze the results. It basically calculates the UV data by convolving a theoretical input sky with a given synthesis beam, then produces a synthesized image. The input submillimeter skies are generated by source count models, which are based on previous observations (see Introduction). The code requires instrument specs as input parameters, including system temperature, array location, and observation period. We use 8 antennas with 4 configurations for SMA and 60 antennas with 3 configurations for ALMA. For each array and configuration, we simulate 256 fields for SMA and 128 for ALMA, and subsequently, estimate their detection and false detection ratios.

## 2.1 Simulating submillimeter skies

The fittings of current observed data are used to produce submillimeter skies. By considering the source counts at bright end, Barger et al. (1999a), Scott et al. (2002) and Borys et al. (2002) give their source counts results in different forms. In the SCUBA 8-mJy survey, Scott et al. (2002) used a source count model to generate full simulated submillimeter skies and use the skies to estimate the incompleteness, positional uncertainties and the effects of confusion in their survey. The equation is shown as the followed

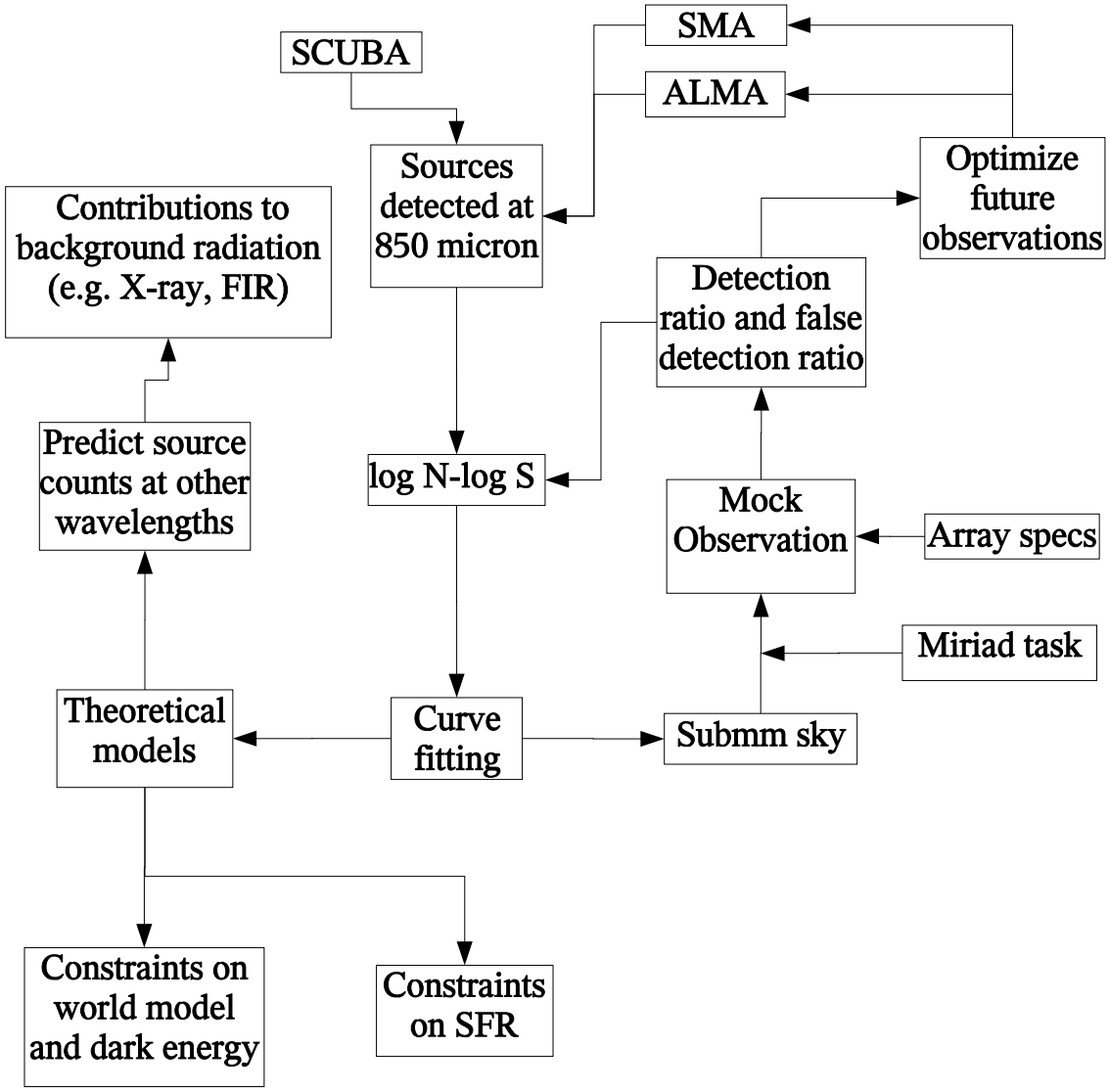


Fig. 2.1 : Flowchart of the relation between each procedure and related sciences in this thesis.

equation.

$$\frac{dN}{dS} = \frac{N_0}{S_0}\{(\frac{S}{S_0})^{\alpha} + (\frac{S}{S_0})^{\beta}\}^{-1} \tag{2.1}$$

where $N_0 = 1.1 \times 10^4$ deg$^{-2}$mJy$^{-1}$, $S_0 = 1.79$ mJy, $\alpha = 0.0$ and $\beta = 3.12$. The source count model is plotted as the dotted curve on Fig. 1.4. The SCUBA scan-map observations toward HDF also give a source count model by Borys et al. (2002),

$$N(> S) = N_0 \left(\frac{S}{S_0}\right)^{-\alpha} \left(1 + \frac{S}{S_0}\right)^{-\beta} \tag{2.2}$$

where $S_0 = 10$ mJy, $\alpha = 0.8$, $\beta = 2.5$ and $N_0 = 1.55 \times 10^4$ deg$^{-2}$.

In this work, we adopt the source count model from Barger et al. (1999). Their observation are reviewed in section 1.4. The source count model is shown as the following equation.

$$N(> S) = \frac{N_0}{a + S^{\alpha}} \tag{2.3}$$

where $N_0$=3.0$\times$10$^4$ deg$^{-2}$mJy$^{-1}$, $\alpha = 3.2$ and $a = 0$ based on $\chi^2$ fitting (Barger et al., 1999). Here we fit all data points from Scott et al. (2002) and Smail et al. (2002) using Eq. 2.3. Our fitting result suggests that $N_0 = 2.03 \pm 0.5 \times 10^4$ deg$^{-2}$mJy$^{-1}$, $\alpha = 2.25 \pm 0.14$ and $a = 1.033 \pm 0.5$ by using least square with unequal weighting. The solid curve in Fig. 2.2 shows the fitting result. The main difference is that the data points we used are more than that of Barger et al. (1999).

The submillimeter sky in which the sources are distributed follows Eq. 2.3 are called Sky 1. Sky 1 predicts that there are no sources at the faint end, so that the number of sources does not increase when the sensitivities of telescopes becoming higher and higher (the detail is shown in Section 4.1). Here we consider two more source count models based on the fitting of the SCUBA data using two power laws. The form is listed as

$$N(> S) = \begin{cases} a_0 S^{\alpha}, & \text{if } S < S_b \\ a_1 S^{\beta}, & \text{if } S \geq S_b. \end{cases} \tag{2.4}$$

The data points brighter than 1 mJy are fitted using the first equation of Eq. 2.4, while the data points fainter than are fitted using the second one. In SCUBA observation,

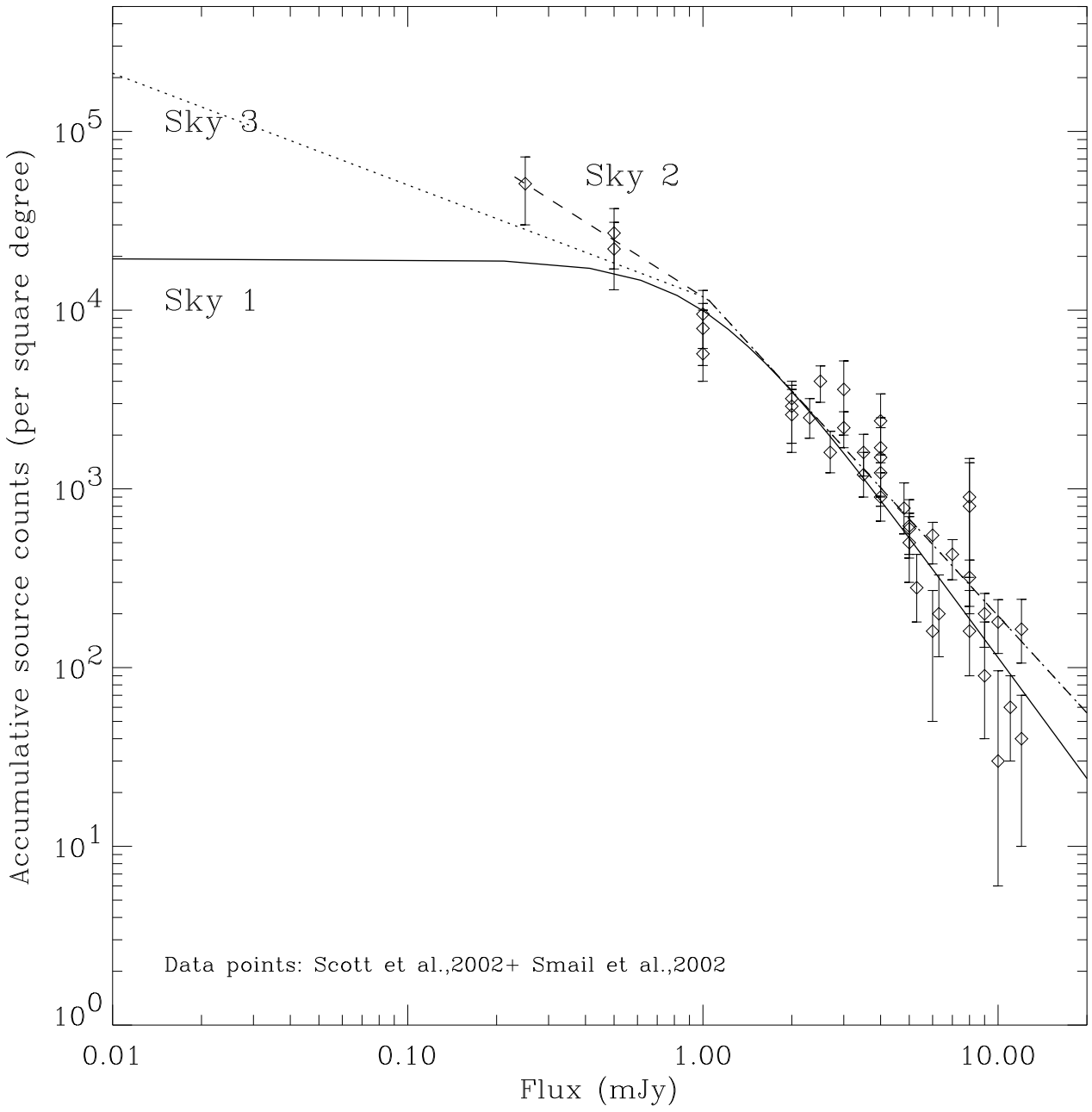


Fig. 2.2 : Three source count models used to generate submillimeter skies in our project.

the data points fainter than 1 mJy are mainly from lensing observations, which are model dependent. The fitting results are

$$N(>S) = \begin{cases} 1.2 \pm 0.8 \times 10^4 S^{-1.05\pm0.71} & \text{if } S < S_b = 1.05 \\ 1.2 \pm 0.2 \times 10^4 S^{-1.80\pm0.10} & \text{if } S \geq S_b = 1.05. \end{cases} \tag{2.5}$$

We calculated the EBL as a constraint for our new source counts. The EBL constrains the slope and the limit flux of source count model, $S_{lim}$. According to COBE FIRAS observation, the 850 $\mu$m EBL measurement gives a background radiation of $4.4\times10^4$ deg$^{-2}$ mJy (Fixsen et al., 1998). Therefore, the total intensity contributed by our source count model must be less or equal to EBL. The following equations shows the evaluation.

$$\begin{aligned} I &= \int_{S_{lim}}^{\infty} S dN \qquad (2.6) \\ &= \int_{S_{lim}}^{1.05} S a_0 \alpha S^{\alpha-1} dS + \int_{1.05}^{\infty} S a_1 \beta S^{\beta-1} dS \\ &= (\alpha a_0) \int_{S_{lim}}^{1.05} S^{\alpha} + \lim_{b\to\infty} \left[ (\beta a_1) \int_{1.05}^{b} S^{\beta} \right] \\ &= \frac{\alpha a_0}{\alpha+1} S^{\alpha+1} \Big]_{S_{lim}}^{1.05} + \lim_{b\to\infty} \left[ \frac{\beta a_1}{\beta+1} S^{\beta+1} \Big]_{1.05}^{b} \right] \\ &= \frac{\alpha a_0}{\alpha+1} \left[ {S_{lim}}^{\alpha+1} - 1.05^{\alpha+1} \right] + 2.7 \times 10^4 \end{aligned}$$

Because of the constraint of EBL, result of Eq. 2.6 is less than EBL. Therefore, we have

$$\begin{aligned} \frac{\alpha a_0}{\alpha+1} \left[ S_{lim}^{\alpha+1} - 1.05^{\alpha+1} \right] + 2.7 \times 10^4 &\leq 4.4 \times 10^4 \quad \text{deg}^{-2} \text{ mJy} \qquad (2.7) \\ \frac{\alpha a_0}{\alpha+1} \left[ {S_{lim}}^{\alpha+1} - 1.05^{\alpha+1} \right] &\leq 1.7 \times 10^4 \\ 2.5 \times 10^5 \left[ S_{lim}^{\alpha+1} - 1.05^{\alpha+1} \right] &\leq 1.7 \times 10^4 \\ \left[ S_{lim}^{\alpha+1} - 1.05^{\alpha+1} \right] &\leq 0.068 \\ {S_{lim}}^{\alpha+1} &\geq 1.146 \\ S_{lim} &\geq 0.28 \qquad \text{mJy} \end{aligned}$$

The $S_{lim}$ should be 0.28 if we assume the submillimeter sources distributed follow Eq. 2.5. This source count model is plotted as dashed curve in Fig. 2.2 as Sky 2.

Eq. 2.5 gives a steep slope for source count curve when the sources are fainter than 1.05. Consider another case, $S_{lim} = 0.01$ mJy, we calculate another source count model.

Assuming power index at faint end of Eq. 2.4 varies between the range $-1.05 \pm 0.71$ and the $S_{lim} = 0.01$. We have the following source count model,

$$N(>S) = \begin{cases} 1.2 \times 10^4 S^{-0.63^{+0.29}_{-1.13}} & \text{if } S < S_b = 1.033 \\ 1.2 \times 10^4 S^{-1.80 \pm 0.10} & \text{if } S \geq S_b = 1.033. \end{cases} \quad (2.8)$$

The source count model gives a gentle slop at faint end and is plotted as dotted curve in Fig. 2.2 as Sky 3. The differences between Sky 1, 2 and 3 are discussed in Section 4.1.

## 2.2 Randomly distributed skies

Since the source counts from models and fittings are determined, we can use these source count models to generate submillimeter skies. The source count models we used in this work are all normalized to 1 square degree. Then, skies are generated by Monte-Carlo method to distribute all points in 1 square degree sky areas. When the 1-square-degree skies are ready, numerous observation regions are selected in each skies by considering the fields of view of SMA and ALMA. For ALMA, its field of view is 18" in diameter at 345 GHz and which of SMA is 36". When the observation field is selected, the pointing positions (e.g. R.A. and Dec.) of the each field is written into a file. So that, we can know the fluxes and positions of the simulated sources, which will be inputs in *Miriad* for mock observations.

There are 2 more conditions when produce a submillimeter sky. The first is that each source is assumed to be the point source without angular size. The second one is that, in this simulations, the effect of antenna attenuation across the primary beam is also considered. We assumed the function of attenuation is a gaussian function. In other words, we will have the 100% flux when the source is at the center of the field of view and have 50% flux when source is at the edge.

Practically, it is not easy to generate the random numbers that follow our desired source counts. In our simulations, we apply the *transformation method* to produce sources fluxes that follow our source count models. The built-in function of IDL provides a random number generator, which generates number between 0 and 1 uniformly. In our case, we have to make the flux distribution of our simulated sources follow the source count model. Therefore, we have to do something more to achieve this requirements. In Fig. 2.3, support that we are going to produce the random numbers between 0 and

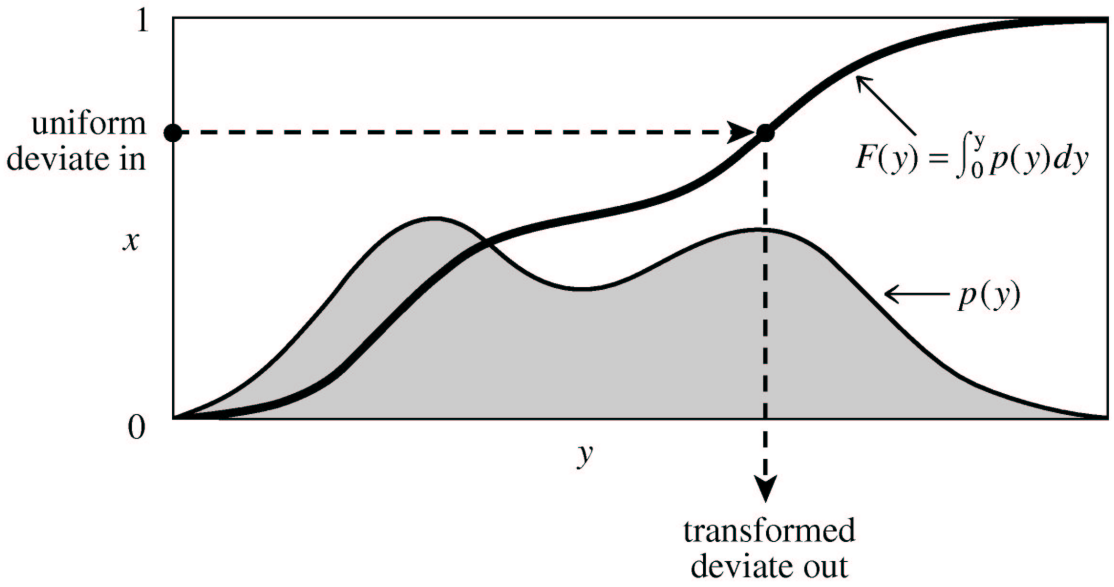


Fig. 2.3 : Illustration for transformation method. (Press et al., 1992)

Y, and its probability follows the distribution function $p(y)$. The integration of $p(y)$ is shown as $\int p(y)$ and the value cumulative probability from 0 to Y should be 1. That can be expressed as

$$F(y) = \int_0^Y p(y)dy = 1. \tag{2.9}$$

From Fig. 2.3, we can understand that $F(y)$ is also a function of $x$, that is $x = F(y)$. The maximum and minimum of $F(y)$ are 1 and 0, respectively. Therefore, we generate the random number between 0 and 1 uniformly as function $F(y)$. Then, take the inverse function of the differential function of $F(y)$, we have the desired probability function $p(y)$.

## 2.3 SMA & ALMA

### 2.3.1 SMA specs

Our mock observations using *Miriad* tasks take many of the instruments' properties into account, so that the specs of the telescopes have to be specified as the input parameters (see Table. 1 and 2) . We estimate the sensitivity of SMA by taking values into Eq. 2.10. There will be 8 antennas in SMA when it is fully functional. The diameter of each antenna is 6 meters. In their design, SMA will have the maximum bandwidth about 2 GHz. SMA will have 6 blocks when it is full functional. The bandwidth of a block is 328 MHz and it contains 4 chunks in each block. Each chunk has 4 chips, each chip can process 64 channels, in maximum. The total channels of SMA correlator is 6144, with bandwidth 1968 MHz and the SMA spectrum resolution can be changed by changing the total number of chips in each chunk.

Assuming there is no phase jitter, so the $n_j$ should be 1. And the correlator quantum efficiency is 0.88 in SMA origin design. The number of orthogonal polarizations correlated is 1 or 2 (2 is for 345 GHz). For the aperture efficiency, it depends on the observation frequency. The aperture efficiencies, $n_a$, are 0.76, 0.71 and 0.51 at 230, 345 and 650 GHz. However, the value is different from the current value. For example, at 230 GHz, the aperture efficiencies in built antennas are below the design value. Here, we take the design value to estimate the array sensitivities at different wavelength. Finally, the last parameter need to be determined is the system temperature. The Fig. 2.4 shows the change of sensitivity with observation time at different wavelength. Here, we take $T_{sys}$=400 K. For 10-hour observations at 230 and 345 GHz, the noise level is down to 1 mJy level. Table. 2.1 is a list of parameters used to estimate the SMA sensitivity.

### 2.3.2 ALMA specs

In the next decade, ALMA will be operated in southern hemi-sphere. It contains 64 antennas, 12 meters in diameter for each antenna, and is located in Llano de Chajnantor, Chile. The range observation frequency of ALMA is from 350 $\mu$m to 10 mm. Also, we estimate the sensitivity level by taking needed values into Eq. 2.10 and Eq. 2.11.

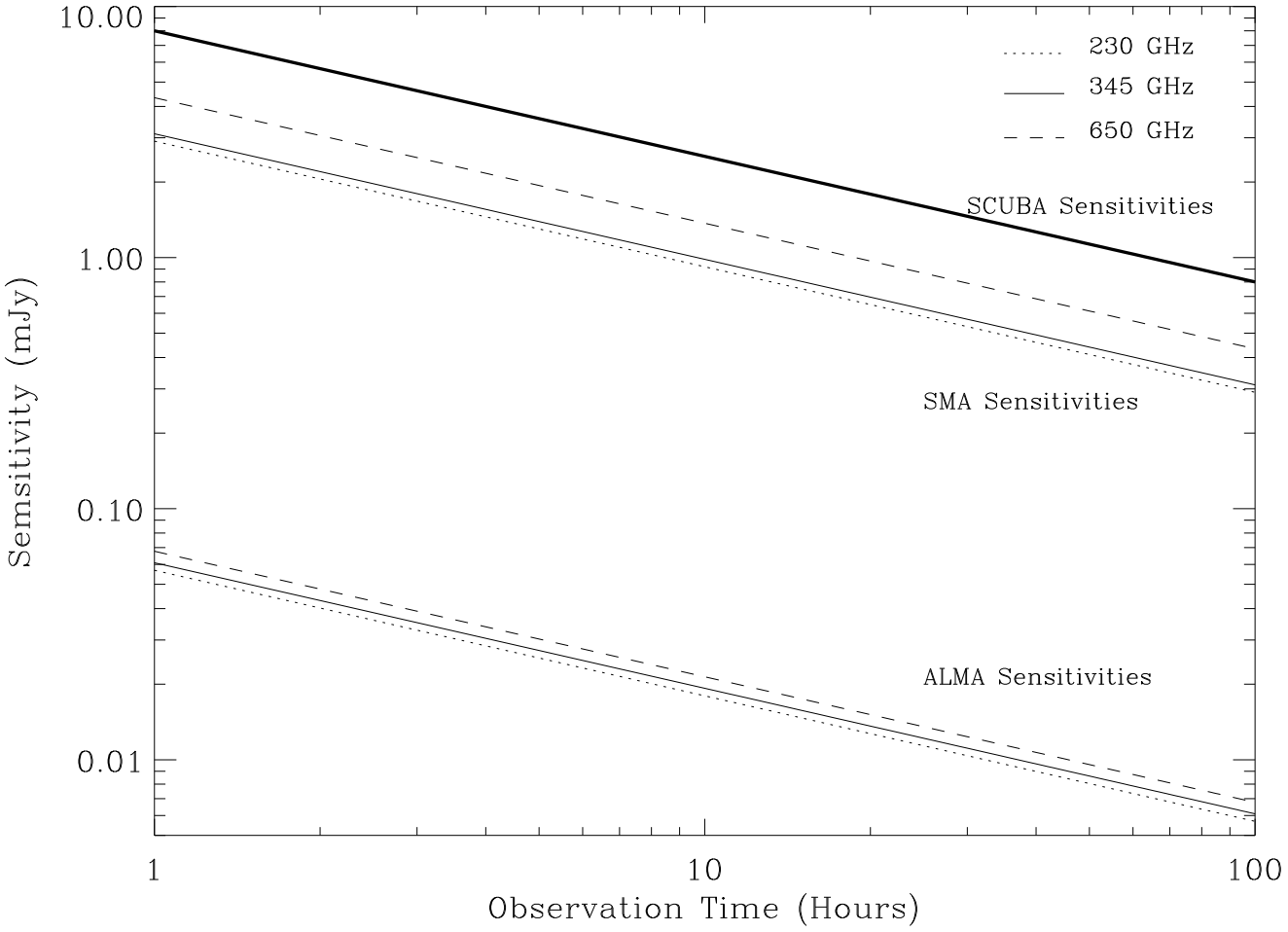


Fig. 2.4 : SMA, ALMA and SCUBA sensitivities. The figures show the relation between integration time and sensitivity of SMA, ALMA and SCUBA. The dotted, solid and dashed line are the sensitivity at 230, 345 and 650 GHz, respectively. The sensitivities of SCUBA are shown as the thick solid curve.

Table 2.1: The SMA parameters used to estimate its sensitivity.(Wilner, 1998)

| Parameter | Value | Notes |
|---|---|---|
| D | 6m | By Design |
| N | 8 | Total antennas of SMA |
| $n_p$ | 1 | 2 (someday) for 345 GHz |
| $\delta\mu$ | 2 GHz | Max bandwidth |
| $n_j$ | 1 | Assuming no phase jitter |
| $n_q$ | 0.88 | Corrector efficiency |
| $n_l$ | 0.9 | Spillover efficiency |
| $n_a$ | 0.76 | Ap. efficiency for 230 GHz |
| | 0.71 | Ap. efficiency for 345 GHz |
| | 0.51 | Ap. efficiency for 650 GHz |

For the correlator quantization efficiency, $n_q$, the designed value of ALMA is 0.95. That is higher than what SMA has. Assuming there is no phase jitter for ALMA, so the $n_j$ should be 1. Taking $n_p = 2$ into Eq. 2.10. In the original design of ALMA, its aperture efficiencies, $n_a$ are 0.75, 0.70 and 0.48 at 230, 345 and 675 GHz. However, the ALMA is not yet contracted, so we have no current value for comparison. The total bandwidth of ALMA is similar to SMA , 2 GHz. The last parameter needed to know is the system temperature. We calculate system temperature using Eq. 2.11. Where we take $n_l = 0.95$ for ALMA. Fig. 2.4 shows the change of sensitivities at different frequencies. The most remarkable feature of ALMA is its higher sensitivity than SMA.

### 2.3.3 Array sensitivity

For a interferometer, its sensitivity can be expressed as the following formula.

$$\Delta S = \frac{2kT_{sys}}{n_j n_q [n_a \pi (D/2)^2][n_p N(N-1)\Delta\nu\Delta t]^{0.5}}, \tag{2.10}$$

where k is the Boltzmann's constant, N is the total amount of antennas in the array, $n_j$ and $n_q$ are the phase jitter efficiency and correlator efficiency, respectively. $\Delta\nu$ is the bandwidth,  is the integration time, $n_p$ is the number of orthogonal polarizations correlated (1 or 2). The term $n_a\pi(D/2)^2$ gives the effective area of an antenna, where

Table 2.2: The ALMA specs.(Bulter, 1999)

| Parameter | Value | Notes |
|---|---|---|
| D | 12m | By Design |
| N | 64 | Total antennas of ALMA |
| $n_p$ | 2 | For all frequencies |
| $\delta\mu$ | 2 GHz | Max bandwidth |
| $n_j$ | 1 | Assuming no phase jitter |
| $n_q$ | 0.95 | Corrector efficiency |
| $n_l$ | 0.95 | Spillover efficiency |
| $n_a$ | 0.75 | Ap. efficiency for 230 GHz |
| | 0.70 | Ap. efficiency for 345 GHz |
| | 0.48 | Ap. efficiency for 650 GHz |

$n_a$ is the aperture efficiency and D is the physical diameter of an antenna. $T_{sys}$ is the system temperature.

As we know, system temperature is depends on the (1) receiver temperature, (2) atmospheric condition (3) the antenna temperature and (4) the cosmic microwave background. The system temperature can be calculated from the following equation.

$$T_{sys} = T_{RX}e^{\tau_0 A} + \eta_l T'_{atm}(e^{\tau_0 A} - 1) + (1 - \eta_l)T'_{sbr}e^{\tau_0 A} + T_{cmb}, \tag{2.11}$$

where, $T_{RX}$ is the receiver temperature, $\tau_0$ is the zenith optical depth, $A$ is airmass, $T'_{atm}$ is the ambient temperature, $T'_{sbr}$ is the temperature for spillover blockage, $\eta_l$ is the (warm) spillover efficiency and $T_{cmb}$ is the cosmic background temperature. Here we take both reliable value into formula. When all parameters are known, the array sensitivity can be calculated as the function of integration time.

### 2.3.4 Array configurations

The array configurations are very important to the synthesis beam. It is an important factor of the interferometer resolution. There are 4 different configurations for SMA and are plotted as Fig. 2.5 [1]. The array configurations are named as A, B, C, D configuration,

[1] http://sma-www.cfa.harvard.edu/private/hawaii/padloc k2.html

from most wide to compact, respectively. Table 2.3.4 shows the pad locations for each configuration.

For ALMA, it will contain 172 pad locations for 29 configurations [2]. Further 42 pads are being optimized for configurations with max baselines between 4 km and 20km and are not yet available. Therefore, in our mock observations, we only use 60 antennas in 29 configurations. Fig. 2.6 shows 4 pad locations for configurations, 1, 11, The configuration 1 is the most compact array and the max baseline is increasing with the configurations.

## 2.4 Miriad tasks for generating synthesis images

We use Miriad as our tool to generate and analyze the mock observation data. Miriad stands for *Multichannel Image Reconstruction, Image Analysis and Display.* This software is used in radio astronomy frequently. We use *Miriad* version 3.1.0 and version 4.0.3 on SPARC and Intel platform, respectively. Version 3.1.0 is released on February 23, 2003 and version 4.0.3 is release on May 14, 2003 from the University of Maryland. We do not *Miriad* older than version 3.1.0 due to its bugs when reducing visibility data from ALMA mock observations.

The first step is to generate the visibility data. Task *uvgen* computes visibility data for a model source distribution at u-v data points. A set of parameters is needed for *uvgen.* The first is *source.* It is a text file contains the sources components, including their flux, positions on the sky and source sizes. The source components are produced using the source count models. The pointing positions of each observation are input to tell where the sources are. Observation time is set in the form of hour ranges. For SMA, the observation time are set to be 50 and 100 hour; while for ALMA, the observation time are 8, 12 and 24 hours with sample interval 0.2 hour. Additionally, the array configurations are also needed to produce visibility data, since they are important for interferometers in calculating UV coverages. For the random noises, we assume that the system temperature is 400 K for both SMA and ALMA. Then, the task *invert* convert visibilities to image. Here we set image size is 512 by 512 pixels. Each pixel represents 0.07 arcsecond in ALMA mock observations and 0.14 arcsecond in SMA observations. This task also produces a dirty beam and a dirty map that will be used by task *clean.*

[2]Dr. John Conway, jconway@oso.chalmers.se

Table 2.3: SMA pad locations and configurations.

| Pad | EW (m) | NS(m) | H (m) | A config. | B config. | C config. | D config. |
|---|---|---|---|---|---|---|---|
| 1 | 0.00 | 0.00 | 0.00 | √ | √ | √ | |
| 2 | -7.47 | -5.84 | -0.01 | | | | √ |
| 3 | -23.76 | -8.23 | -0.06 | | | | √ |
| 4 | -25.16 | 1.37 | -0.04 | | | | √ |
| 5 | -18.98 | 16.62 | -0.08 | | | | √ |
| 6 | -10.11 | 13.01 | -0.01 | | | | √ |
| 7 | -20.09 | -15.73 | -0.10 | | | √ | √ |
| 8 | -63.88 | -22.06 | -3.21 | | | √ | |
| 9 | -68.02 | 5.59 | -4.79 | | | √ | |
| 10 | -51.03 | 44.88 | -6.68 | | | √ | |
| 11 | -27.03 | 34.55 | -5.13 | | √ | | |
| 12 | -53.67 | -42.30 | -2.75 | √ | √ | √ | |
| 13 | 180.98 | -58.59 | -11.48 | √ | √ | √ | |
| 14 | 178.82 | 3.36 | -11.45 | | √ | | |
| 15 | 136.76 | 118.97 | -11.77 | | √ | | |
| 16 | -71.55 | 95.53 | -10.84 | | √ | | |
| 17 | -133.60 | -107.01 | -10.38 | | √ | | |
| 18 | -349.68 | 42.55 | -24.69 | √ | | | |
| 19 | -479.03 | 67.96 | -36.17 | √ | | | |
| 20 | -251.59 | 335.61 | -18.58 | √ | | | |
| 21 | -73.52 | 321.22 | -20.29 | √ | | | |
| 22 | -23.73 | 9.20 | -0.03 | | | | √ |
| 23 | -59.57 | 34.35 | -6.64 | | | √ | |
| 24 | -169.88 | 109.69 | -11.73 | | | | |

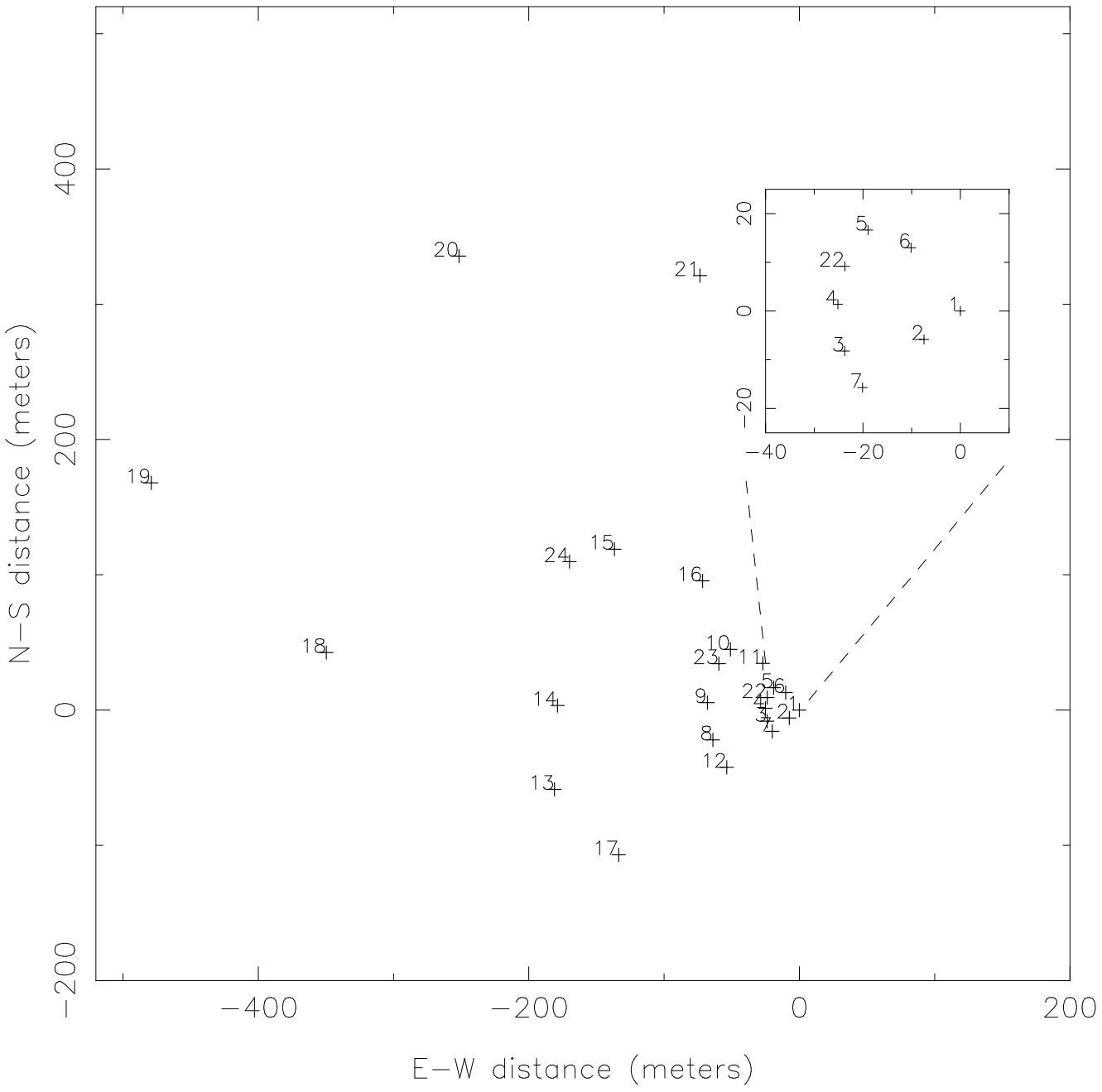


Fig. 2.5 : The SMA pad locations.

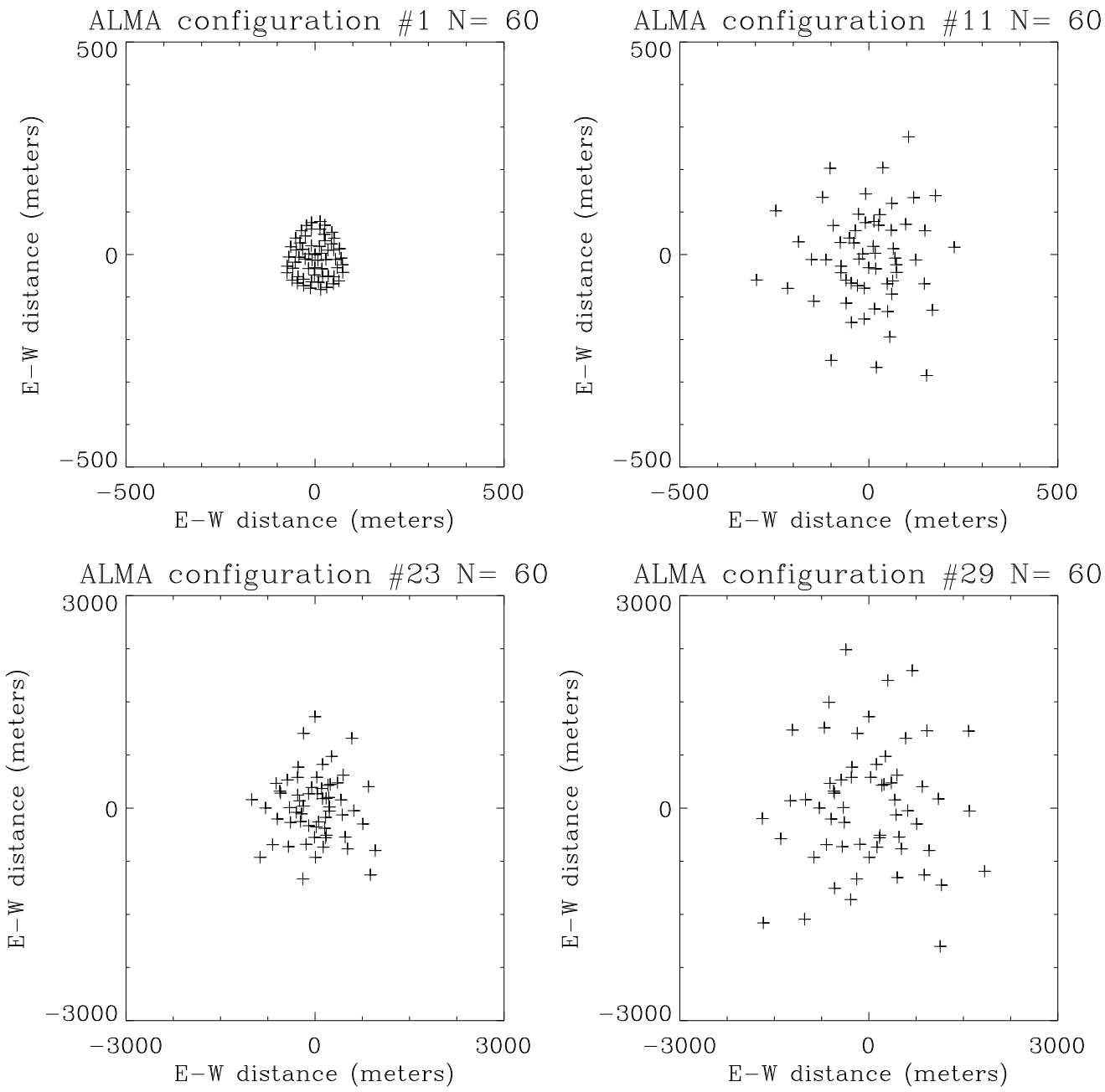


Fig. 2.6 : The pad locations of 4 configurations of ALMA.

Task *clean* and *restor* are used to produce the synthesis map. The task *clean* deconvolve the dirty map with dirty beam and produce a clean component. After convolving the clean component with a Gaussian beam, task *restor* produces a clean map.

## 2.5 Data Analysis

After all *Miriad* tasks, we use IDL programs to find out sources in the maps. The main task used is *find.pro* from *astrolib*[3]. It is adopted from the algorithm *DAOPHOT*, which is used widely in finding point sources in optical photometry. By giving the detection threshold, beam size, sharpness and roundness threshold for the sources, it returns the flux, position sharpness and roundness of sources found in a giving image. In this section, we describe the detail settings of source finding procedure.

The *Miriad* tasks produce maps in FITS format and PostScript (PS) format with dimension 512 by 512 pixels. For SMA, each pixel corresponds to 0.14 arcsecond. And the pixel scale is 0.07 arcsecond per pixel for ALMA. The maps in PS format are used to check the maps produced in IDL analysis. While images in FITS format are input into IDL program. In IDL program, the first step is to check the fluxes and positions of input sources. The source with flux larger than given significance level are plotted as a square mark. Or the source are plotted as cross marks.

Unlike optical image, the shape of point source in synthesis image is decided by synthesis beam, which is depending on the configuration and observation time (in other word, it is uv coverage). Therefore, the sources in our mock observation image are usually elongated. We use extend the thresholds for finding point sources. For each mock observation image, we use *find.pro* program with different detection threshold. The threshold is give in the form of significance level, from 2.0 to 5.0$\sigma$. The *DAOPHOT* algorithm searches sources not only depending on the significance levels, but also sharpness and roundness. Here we set the upper and lower limit of sharpness as 0.07 and 1.0, respectively. There are no limits of roundness are set for detecting. The strategies described above do help us detect those largely elongated sources. However, the disadvantage is that the program found more fake sources when the threshold ranges become larger and larger. Those fake sources confuse our detections. Therefore, we apply a calibration

[3]http://idlastro.gsfc.nasa.gov/homepage.html

to reject those fake sources. For each mock observation image, we produce 2 more images, which contains background only without input any mock sources. Then, the same parameters are applied onto mock image and those 2 background images to detecting sources. So that, the sources detected on those background images are all fake sources. By comparing their position and flux, we can reject the fake sources in mock observation image. In practice, each source in background image are compared with the source in mock observation image. If difference in position is within 2 arc-seconds and difference in flux is less than $3\sigma$, we regard these two points are fake sources. By applying this procedure, we can reducing the false detection our mock observation images. In next chapter, our results of SMA and ALMA simulations are shown.