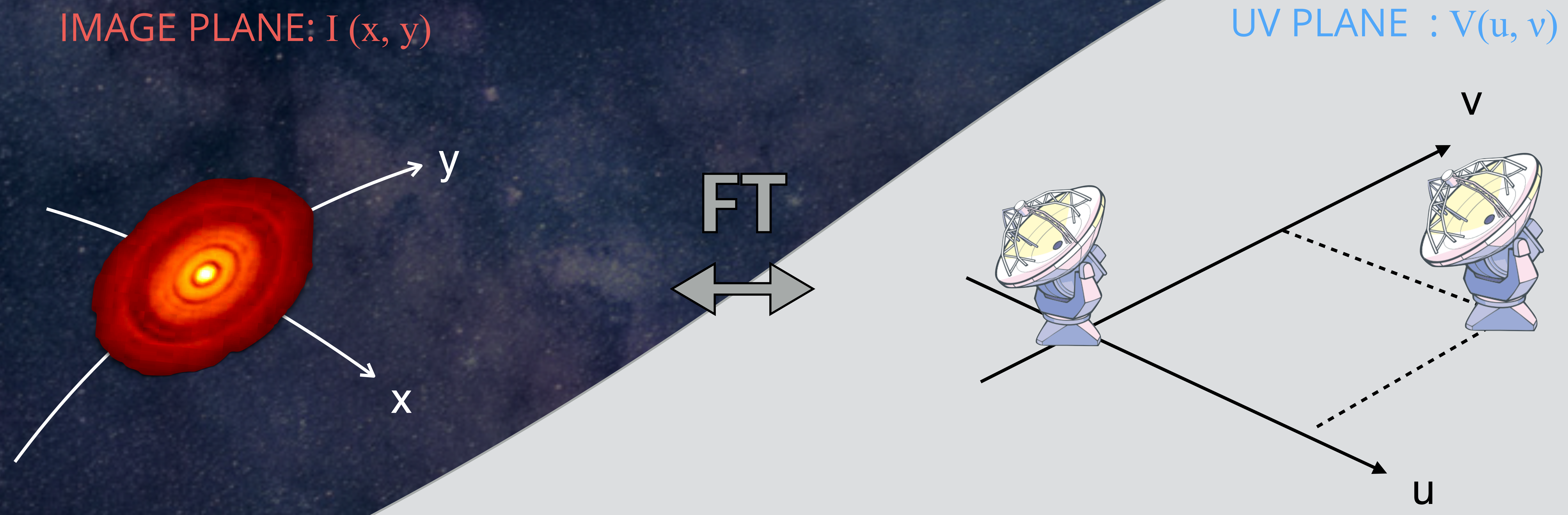
ALMA Imaging Workshop

Self-calibration (for improving an image sensitivity)

Masayuki Yamaguchi
ASIAA, Taiwan

ALMA Imaging workshop, February 14 2025

Observation equation for radio interferometry



$$V(u, v) = \iint I(x, y) e^{-2\pi i(uy+vx)} dxdy$$

Background : The observed visibility is distorted by various factors

Visibility: $V^{\text{true}}(t) = \int I(x,y)e^{i(ux+vy)}dydy$

Ideally: $V^{\text{obs}}(u,x) = V^{\text{true}}(u,v)$

But, in practice

$V^{\text{obs}}(u,x) \neq V^{\text{true}}(u,v)$ and $V^{\text{obs}}(u,x) = \mathcal{G}V^{\text{true}}(u,v)$

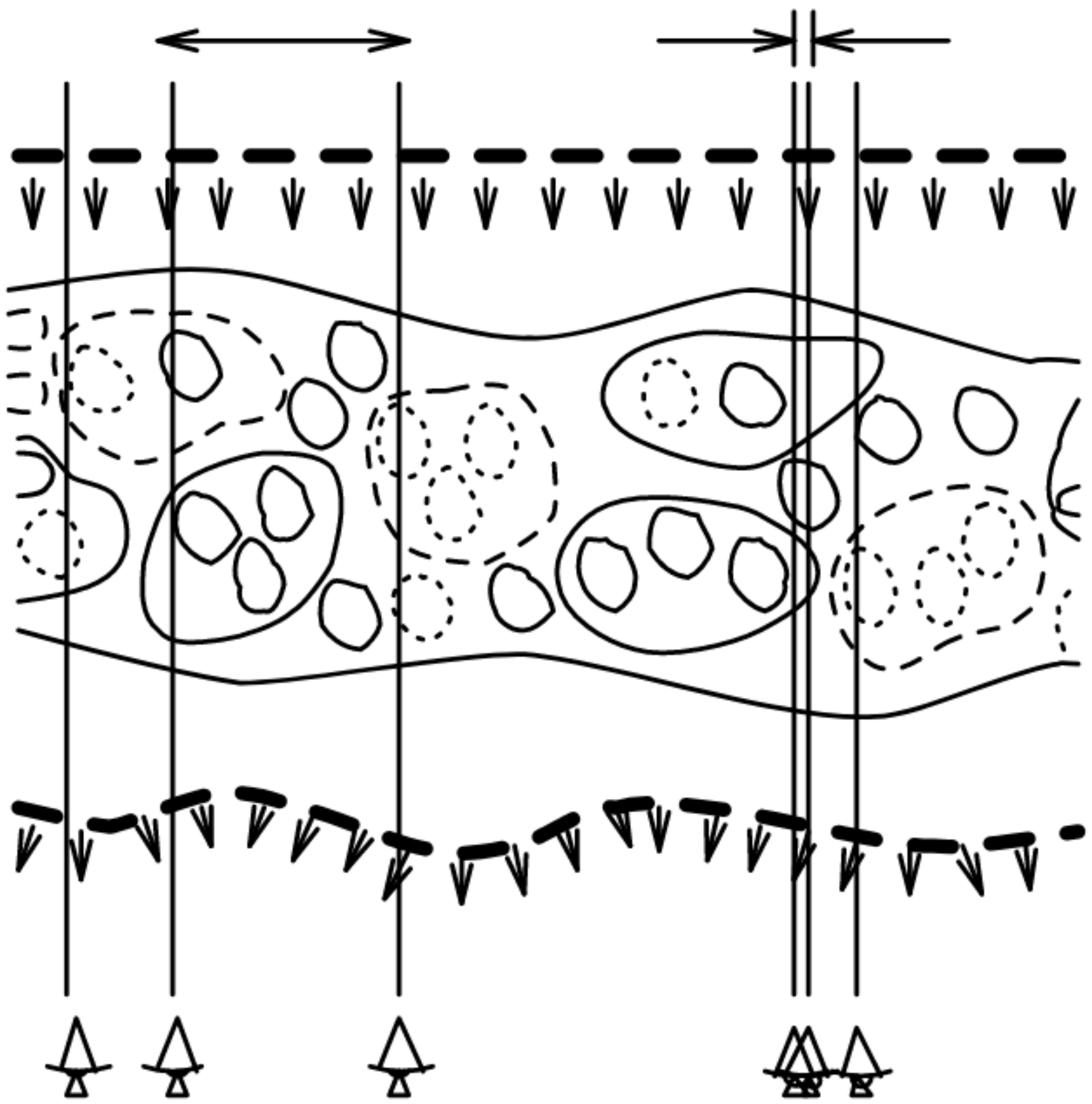
We suffer from residual gain error $\mathcal{G}_{ij}(t)$ even after performing the data calibration (flux, bandpass, phase, etc..).

Atmospheric attenuation $G_{atm}(t)$

Antenna $G_{ant}(t)$

Receiver stability $G_{rc}(t)$

$$\mathcal{G}_{ij}(t) = G_{ij}^{\text{atm}}(t)G_{ij}^{\text{ant}}(t)G_{ij}^{\text{rc}}(t)\cdots$$



Carilli et al 1999
(from Synthesis Imaging in Radio Astronomy II)

Background : The observed visibility is distorted by various factors

So what we want to do is to estimate the complex gains and obtain the true visibility.

$$V^{\text{corr}} = \frac{V^{\text{obs}}}{\mathscr{G}} = \frac{\cancel{\mathscr{G}} V^{\text{true}}}{\cancel{\mathscr{G}}} = V^{\text{true}}$$

(selfcal data)

Self-calibration is the approach to solve the gains using χ^2 calculation between the observation and a custom model.

The principle of self-calibration

The relationship between the observed visibility $V_{ij}^{\mathrm{obs}}(t)$ on the i-j baseline and the true visibility $V_{i,j}^{\mathrm{true}}(t)$ can be described generally as

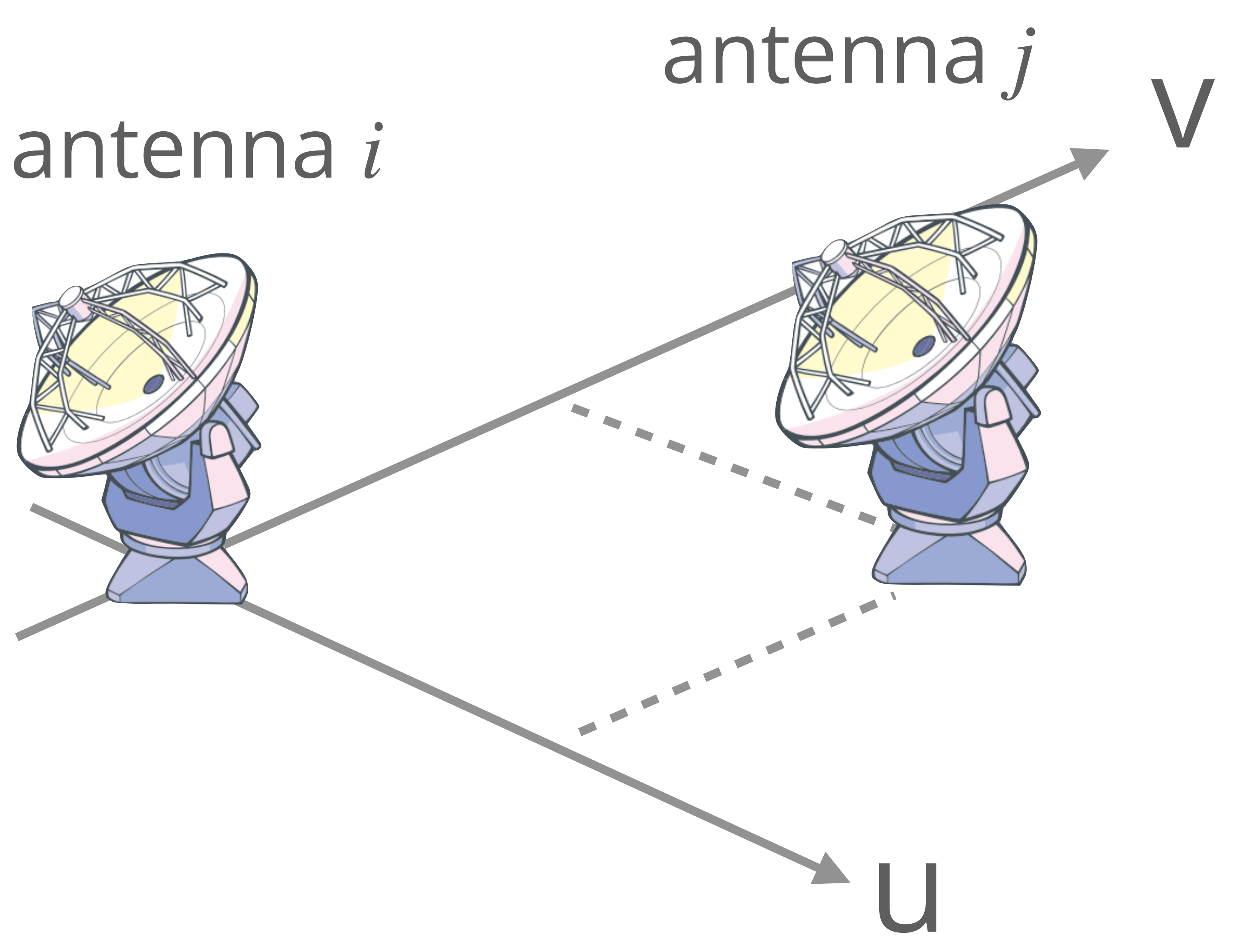
$$V_{ij}^{\mathrm{obs}}(t) = \mathcal{G}_{ij}(t) V_{ij}^{\mathrm{true}}(t) \, .$$

The baseline-based complex gain $\mathcal{G}_{ij}(t)$ can be approximated by the product of the two associated antenna-based complex gain $g_i(t)$ and $g_j^*(t)$;

amp phase

$$\mathcal{G}_{ij}(t) = g_i(t) g_j^*(t) = a_i(t) a_j(t) e^{i\left\{\phi_i(t) - \phi_j(t)\right\}},$$

where $a_i(t)$ is an antenna-based amplitude correction and $\phi_i(t)$ is the antenna-based phase correction.



The principle of self-calibration

Here we assume that the model V^{model} should be astronomically plausible.
The sum of squares of residuals $\mathcal{S}$ can be described as

free parameters

$$\mathcal{S} = \sum_{k} \sum_{\substack{i,j \\ i \neq j}} w_{ij}\left(t_k\right) \left| V_{ij}^{\text{obs}}\left(t_k\right) - \underline{\mathcal{G}_{ij}(t)} V_{ij}^{\text{model}}\left(t_k\right) \right|^2$$

$$= \sum_{k} \sum_{\substack{i,j \\ i \neq j}} w_{ij}\left(t_k\right) \left| V_{ij}^{\text{model}}\left(t_k\right) \right|^2 \left| X_{ij}\left(t_k\right) - \mathcal{G}_{ij}(t) \right|^2$$

~ 1

$$= \sum_{k} \sum_{\substack{i,j \\ i \neq j}} w_{ij}\left(t_k\right) \left| V_{ij}^{\text{model}}\left(t_k\right) \right|^2 \left| X_{ij}\left(t_k\right) - \mathcal{G}_{ij}(t) \right|^2 \quad \text{where} \quad X_{ij}(t) = \frac{V_{ij}^{\text{obs}}(t)}{V_{ij}^{\text{model}}(t)}.$$

In this case, $X_{ij}(t)$ would have around a unity over the baseline.

The χ^2 then solves for the complex gains $\mathcal{G}_{ij}(t)$ when minimizing $\mathcal{S}$.

The principle of self-calibration

The χ^2 solves for the complex gains $\mathcal{G}_{ij}(t)$ when minimizing $\mathcal{S}$.
We then get the complex gains $\mathcal{G}_{ij}(t)$.

Finally, the corrected (self-calibrated) visibility $V_{ij}^{\mathrm{corr}}(t)$ can be obtained as

selfcal visibility

$$V_{ij}^{\mathrm{corr}}(t) = \frac{V_{ij}^{\mathrm{obs}}(t)}{\mathcal{G}_{ij}(t)} = \frac{\cancel{\mathcal{G}_{ij}(t)} V_{ij}^{\mathrm{true}}(t)}{\cancel{\mathcal{G}_{ij}(t)}} = V_{ij}^{\mathrm{true}}(t)$$

➡ In practice, the gain cannot be solved in this single calculation.
We usually do iterative calculations to get the true visibility.

➡ Next to "iterative self-calibration" approach

Workflow of iterative self-calibration

① Make a model image of the target source ➡ We employ "CLEAN model"

V^{obs} → CLEAN imaging → I^{model}

② Convert the model image into the model visibility

I^{model} → FT → V^{model}

③ Solve for the complex gains (phase or/and amp) using χ^2 calculation.

$$\mathcal{S} = \sum_{k} \sum_{\substack{i,j \\ i \neq j}} w_{ij}\left(t_k\right) \left| V_{ij}^{\mathrm{obs}}\left(t_k\right) - \mathcal{G}_{ij}(t) V_{ij}^{\mathrm{model}}\left(t_k\right) \right|^2$$

④ Obtain the corrected (or self-calibrated) visibility using the solved gains.

$$V_{ij}^{\mathrm{corr}}(t) = \frac{V_{ij}^{\mathrm{obs}}(t)}{\mathcal{G}_{ij}(t)}$$

⑤ Repeat to (1) - (4) steps, until you are satisfied with SNR on the resulting image.

Self-calibration with CASA

Workflow of iterative self-calibration with CASA

observed visibility V^{obs}

CLEAN imaging:
generate CLEAN image
and CLEAN model image (vis: V^{model})

V^{obs} (MS-file)
contains "three columns"

data	corrected	model
observations		

used for
CLEAN imaging

Workflow of iterative self-calibration with CASA

observed visibility V^{obs}

CLEAN imaging:
generate CLEAN image
and CLEAN model image (vis: V^{model})

V^{obs} (MS-file)
contains "three columns"

data	corrected	model
observations		CLEAN model

After CLEAN imaging

used for
CLEAN imaging

Workflow of iterative self-calibration with CASA

observed visibility V^{obs}

↓

CLEAN imaging:
generate CLEAN image
and CLEAN model image (vis: V^{model})

↓

“gaincal” task:
solve gain errors $\mathscr{G}$ using
χ^2 calculation between V^{obs} and V^{model}

V^{obs} (MS-file)
contains “three columns”

data	corrected	model
observations		CLEAN model

After CLEAN imaging

data	corrected	model
observations		CLEAN model

χ^2

used for
CLEAN imaging

Workflow of iterative self-calibration with CASA

observed visibility V^{obs}

CLEAN imaging:
generate CLEAN image
and CLEAN model image (vis: V^{model})

"gaincal" task:
solve gain errors $\mathscr{G}$ using
χ^2 calculation between V^{obs} and V^{model}

"applycal" task:
apply the gain table to V^{obs}

V^{obs} (MS-file)
contains "three columns"

data	corrected	model
observations		CLEAN model

After CLEAN imaging

data	corrected	model
observations		CLEAN model

χ^2

data	corrected	model
observations	self-cal data	CLEAN model

used for
CLEAN imaging

Workflow of iterative self-calibration with CASA

observed visibility V^{obs}

CLEAN imaging:
generate CLEAN image
and CLEAN model image (vis: V^{model})

"gaincal" task:
solve gain errors $\mathscr{G}$ using
χ^2 calculation between V^{obs} and V^{model}

"applycal" task:
apply the gain table to V^{obs}

"split" task: overwrite data column

V^{obs} (MS-file)
contains "three columns"

data	corrected	model
observations		CLEAN model

After CLEAN imaging

data	corrected	model
observations		CLEAN model

χ^2

data	corrected	model
observations	self-cal data	CLEAN model
self-cal data		CLEAN model

used for
CLEAN imaging

Workflow of iterative self-calibration with CASA

observed visibility V^{obs}

↓

CLEAN imaging:
generate CLEAN image
and CLEAN model image (vis: V^{model})

↓

“gaincal” task:
solve gain errors $\mathscr{G}$ using
χ^2 calculation between V^{obs} and V^{model}

↓

“applycal” task:
apply the gain table to V^{obs}

↓

“split” task: overwrite data column

↓

CLEAN imaging again:
generate CLEAN image with selfcal data V^{corr}

V^{obs} (MS-file)
contains “three columns”

data	corrected	model
observations		CLEAN model

After CLEAN imaging

data	corrected	model
observations		CLEAN model

χ^2

data	corrected	model
observations	self-cal data	CLEAN model
self-cal data		CLEAN model

used for
CLEAN imaging

Workflow of iterative self-calibration with CASA

observed visibility V^{obs}

↓

CLEAN imaging:
generate CLEAN image
and CLEAN model image (vis: V^{model})

↓

"gaincal" task:
solve gain errors $\mathscr{G}$ using
χ^2 calculation between V^{obs} and V^{model}

↓

"applycal" task:
apply the gain table to V^{obs}

↓

"split" task: overwrite data column

↓

CLEAN imaging again:
generate CLEAN image with selfcal data V^{corr}

The number of iterations of self-calibration

Three rounds of phase calibrations (down to the integration time) and one round of amp & phase.
→ five times in total.

V^{obs} (MS-file)
contains "three columns"

data	corrected	model
observations		CLEAN model

After CLEAN imaging

data	corrected	model
observations		CLEAN model

χ^2

data	corrected	model
observations	self-cal data	CLEAN model
self-cal data		CLEAN model

used for
CLEAN imaging

Workflow of iterative self-calibration with CASA

observed visibility V^{obs}

↓

CLEAN imaging:
generate CLEAN image
and CLEAN model image (vis: V^{model})

↓

"gaincal" task:
solve gain errors $\mathcal{G}$ using
χ^2 calculation between V^{obs} and V^{model}

↓

"applycal" task:
apply the gain table to V^{obs}

↓

"split" task: overwrite data column

↓

CLEAN imaging again:
generate CLEAN image with selfcal data V^{corr}

↓

obtain selfcal V^{corr} and final CLEAN image

The number of
iterations of self-calibration

Three rounds of phase calibrations (down to the integration time) and one round of amp & phase.
→ five times in total.

V^{obs} (MS-file)
contains "three columns"

data	corrected	model
observations		CLEAN model

After CLEAN imaging

data	corrected	model
observations		CLEAN model

χ^2

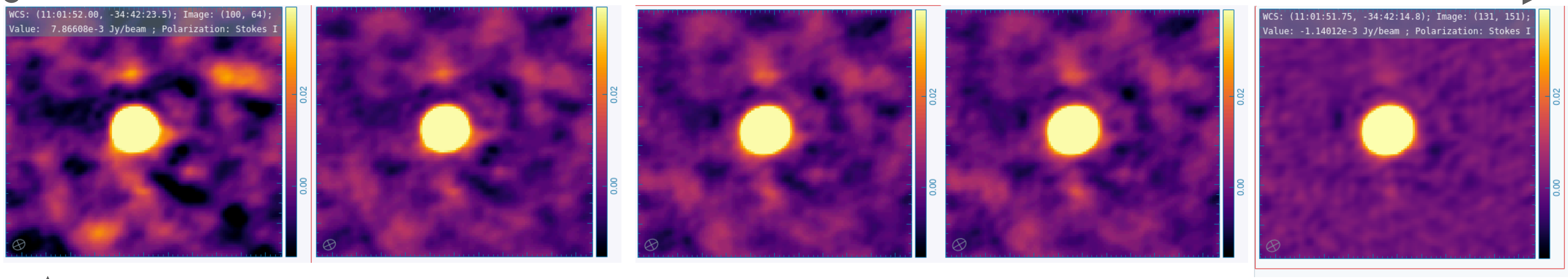
data	corrected	model
observations	self-cal data	CLEAN model
self-cal data		CLEAN model

used for
CLEAN imaging

Resulting images after the self-calibrations

CARTA allows you to check the resulting images and calculate their SNR (peak intensity/RMS noise).

1st CLEAN image	2nd CLEAN image	3rd CLEAN image	4th CLEAN image	5th CLEAN image
without selfcal	after first phase cal solint = “inf” (entire scan duration)	after second phase cal solint = “170sec” (half scan duration)	after third phase cal solint=“30sec” (1/10 scan duration)	after amp & phase cal solint=“inf” (entire scan duration)



Hands-on session

ALMA data information

We use continuum data ("**twhya_ALMAb7_cont.ms**") for the tutorial.

Target source : the protoplanetary disk "TW Hydra"
observed at Band 7 with ALMA 12m array (Project 2011.0.00340.S).

The data information details in First Look at Imaging CASA 6.

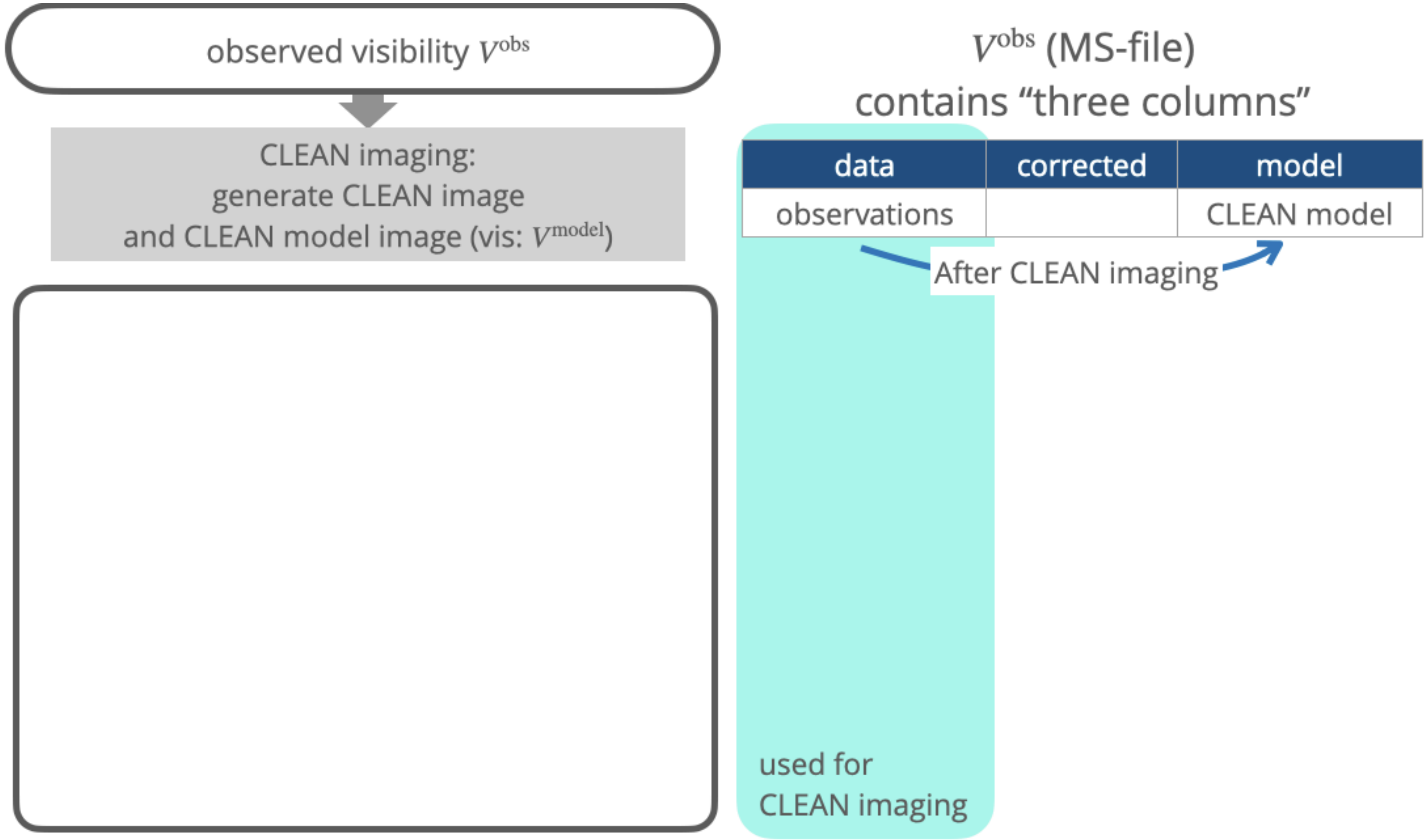
With "listobs" task in CASA: listobs(vis = "**twhya_ALMAb7_cont.ms**", listfile = "**twhya_ALMAb7_cont.ms.listfile**")

```
Data records: 53161       Total elapsed time = 4268.11 seconds
  Observed from   19-Nov-2012/07:56:23.5   to   19-Nov-2012/09:07:31.6 (UTC)

Fields: 1
  ID   Code Name            RA              Decl           Epoch   SrcId     nRows
  0    none TW Hya          11:01:51.796000 -34.42.17.36600 J2000   0         53161

Spectral Windows:  (1 unique spectral windows and 1 unique polarization setups)
  SpwID  Name                           #Chans   Frame   Ch0(MHz)  ChanWid(kHz)  TotBW(kHz) CtrFreq(MHz) BBC Num  Corrs
  0      ALMA_RB_07#BB_2#SW-01#FULL_RES    16   TOPO  372540.105    14648.438   234375.0 372649.9688       2  XX  YY
```

Hands-on session: look at "selfcal.py"

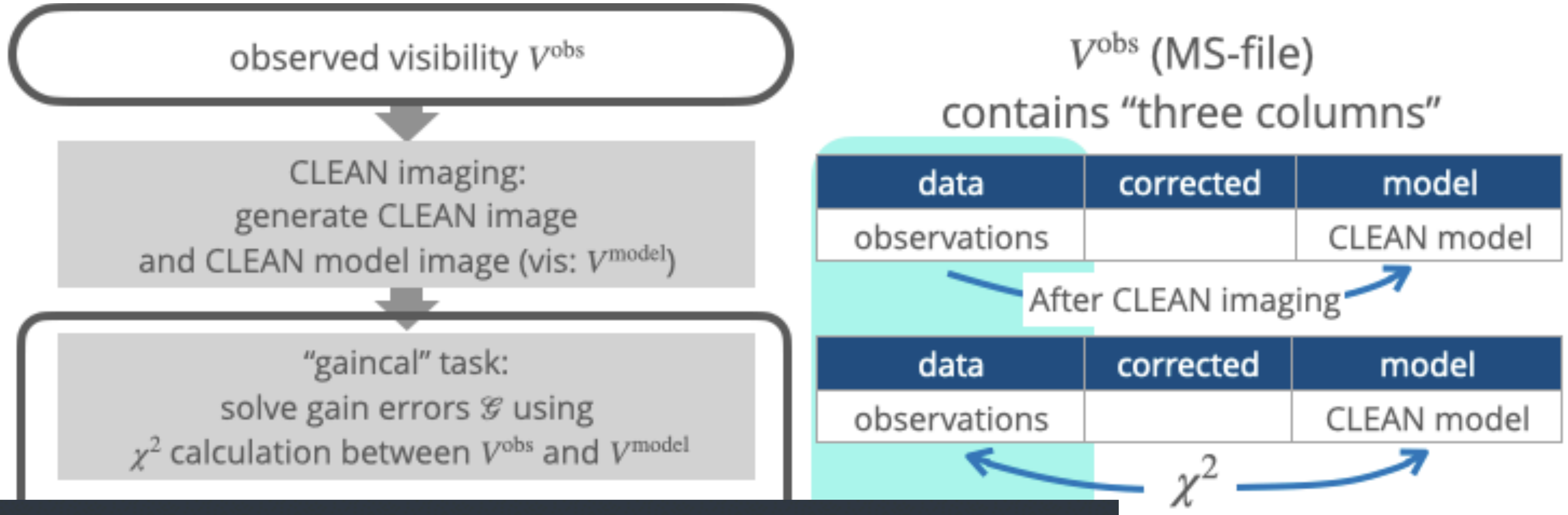
1. Executing CLEAN imaging with the observed visibility (MS-file)
2. After the imaging, the CLEAN model visibility is automatically stored in the model column of the MS-file

```
######################
# first round
######################

# t-clean imaging

os.system('rm -rf first_clean.*')
tclean(vis='twhya_ALMAb7_cont.ms',
       imagename='first_clean',
       datacolumn  ='data',
       field='0',
       spw='',
       specmode='mfs',
       deconvolver='hogbom',
       nterms=1,
       gridder='standard',
       mask = 'circle[[125pix,125pix],3.0arcsec]',
       imsize=[250,250],
       cell=['0.1arcsec'],
       weighting='natural',
       threshold='0mJy',
       niter=10000,
       interactive=False,
       savemodel='modelcolumn')
```



Hands-on session

```
os.system("rm -rf phase.cal")
gaincal(vis='twhya_ALMAb7_cont.ms',
        caltable="phase.cal",
        field="0",
        spw  = '*',
        solint="inf",
        calmode="p",
        refant="DV22",
        gaintype="G")
```



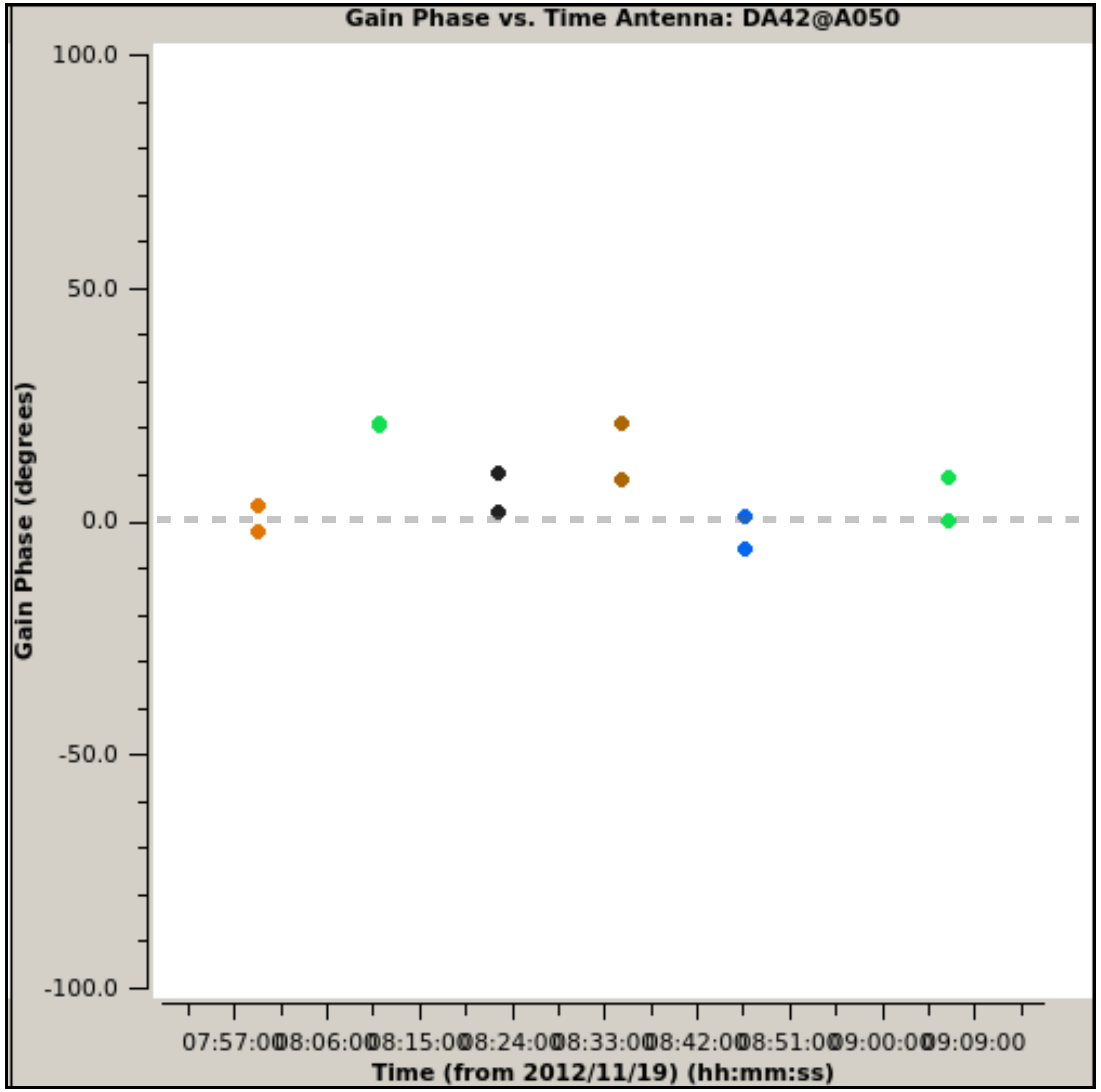

```
# gaincal -- Determine temporal gains from calibrator observations
vis                = ''       # Name of input visibility file
caltable           = ''       # Name of output gain calibration table
field              = ''       # Select field using field id(s) or field name(s)
spw                = ''       # Select spectral window/channels
solint             = 'inf'    # Solution interval
refant             = ''       # Reference antenna name(s)
solnorm            = False    # Normalize (squared) solution
gaintype           = 'G'      # amplitudes (G, T only) Type of gain solution
calmode            = 'p'      # Type of solution ('ap':amp&phase, 'p':phase, 'a':amplitude)
```

If you set **solint = 'inf'**, then the solution interval (time) will be the entire scan duration for each scan.
gaintype='G' solves for separate gain solutions for XX and YY, allowing for polarization-dependent phase corrections.
gaintype='T' averages gains over both polarizations, making it more stable and less noisy.

Hands-on session

After performing the gaincal, let's execute "plotms" task to see the phase gain solutions.

```
plotms(vis='phase.cal',
       xaxis='time',
       yaxis='phase',
       iteraxis='antenna',
       gridrows=5,
       gridcols=6,
       coloraxis='scan',
       plotrange=[0,0,-100,100],
       showgui     = False,
       overwrite   = True,
       height = 2500,
       width = 3000,
       plotfile    = 'phasecal.png')
```

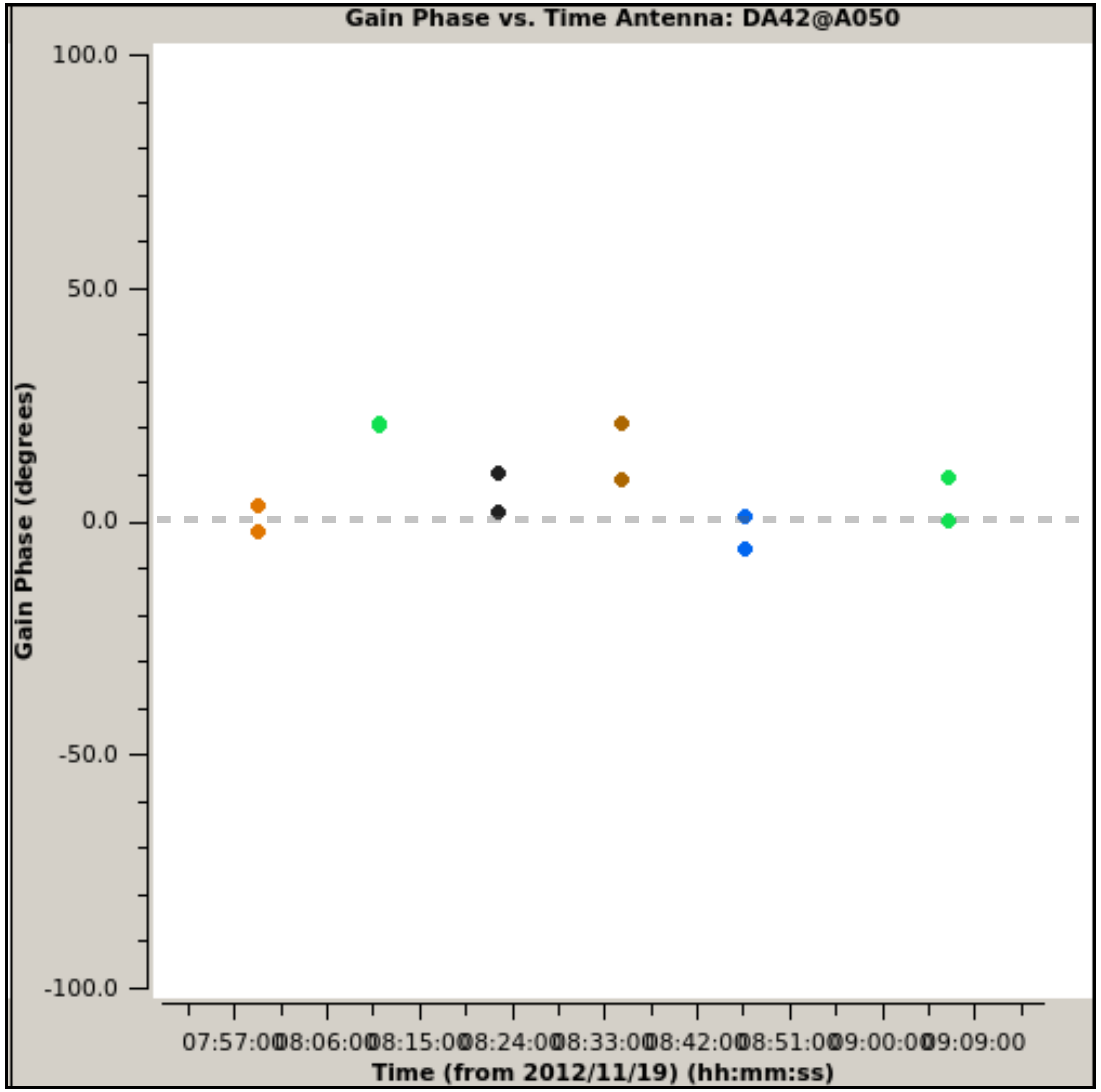


Hands-on session

1. The Figure on the left shows the gain phase after the solving of the complex gains with gaincal (solint = ‘inf').
2. If the phase points are stable around zero, it means that self-calibration successfully corrected phase errors.
3. The ALMA data contain dual polarization (XX and YY). The gain solutions for phase calibration must be considered separately for each polarization.
4. Since this records both XX and YY, we see two phase solutions per time step in the figure.

$$g_i^{XX}(t) = a_i^{XX}(t)e^{i\phi_i^{XX}(t)}$$

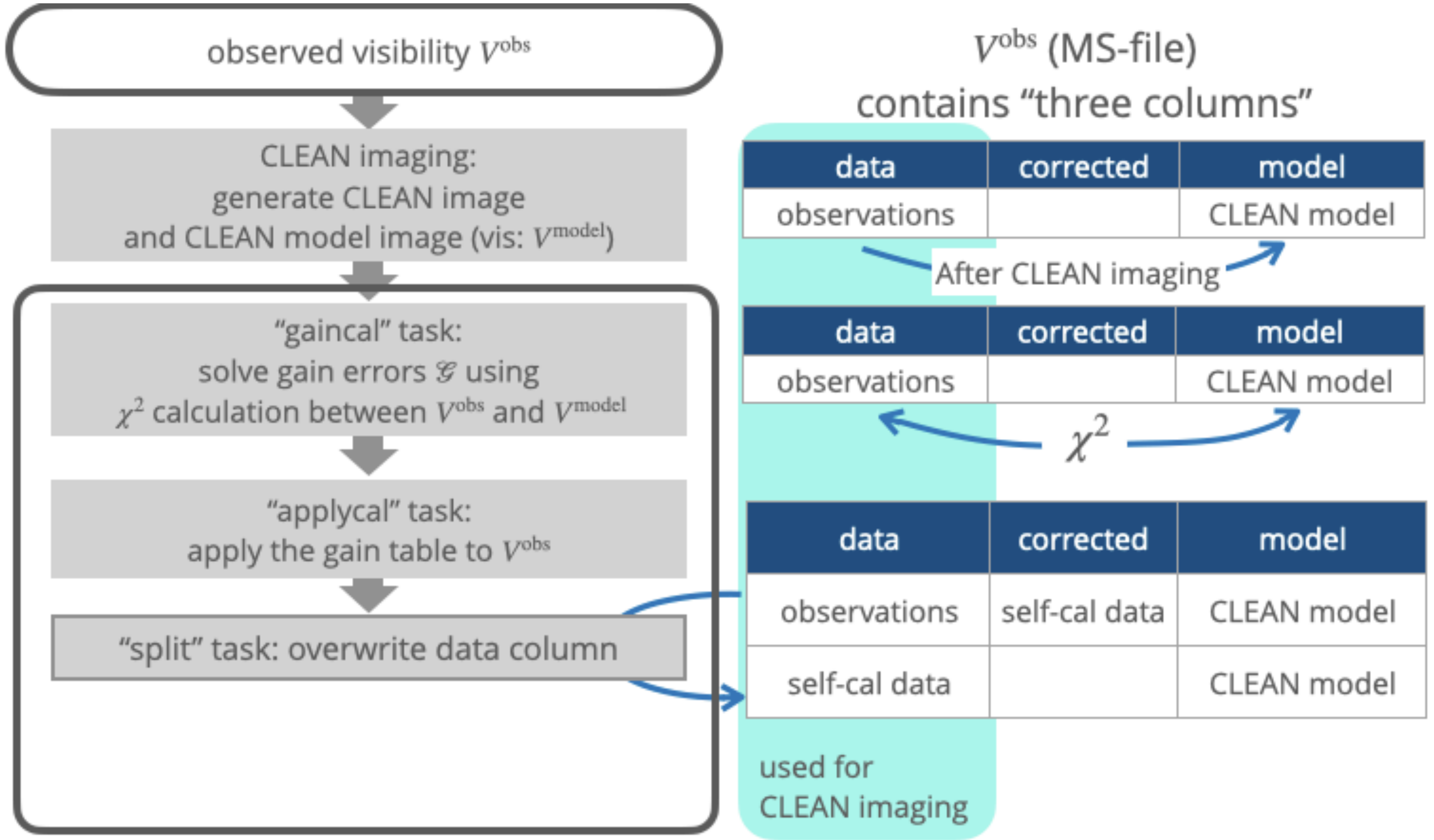
$$g_i^{YY}(t) = a_i^{YY}(t)e^{i\phi_i^{YY}(t)}$$



Hands-on session

1. "applycal" task applies the gain table ("phase.cal") to the observations to correct the gain errors and stores the self-calibrated data in the corrected columns.
2. "split" task overwrites "observation data" with "selfcal data" in the data column.

```
applycal(vis="twhya_ALMAb7_cont.ms",
         calwt = True,
         gaintable=["phase.cal"],
         interp="linear",
         )

os.system("rm -rf twhya_ALMAb7_cont_selfcal.ms*")
split(vis="twhya_ALMAb7_cont.ms",
      outputvis="twhya_ALMAb7_cont_selfcal.ms",
      datacolumn="corrected")
```



Hands-on session

The number of iterations of self-calibration

Three rounds of phase calibrations (down to the integration time) and one round of amp & phase.
→ five times in total.

observed visibility V^{obs}

↓

CLEAN imaging:
generate CLEAN image
and CLEAN model image (vis: V^{model})

↓

"gaincal" task:
solve gain errors $\mathscr{G}$ using
χ^2 calculation between V^{obs} and V^{model}

↓

"applycal" task:
apply the gain table to V^{obs}

↓

"split" task: overwrite data column

↓

CLEAN imaging again:
generate CLEAN image with selfcal data V^{corr}

↓

obtain selfcal V^{corr} and final CLEAN image

V^{obs} (MS-file)
contains "three columns"

data	corrected	model
observations		CLEAN model

After CLEAN imaging

data	corrected	model
observations		CLEAN model

χ^2

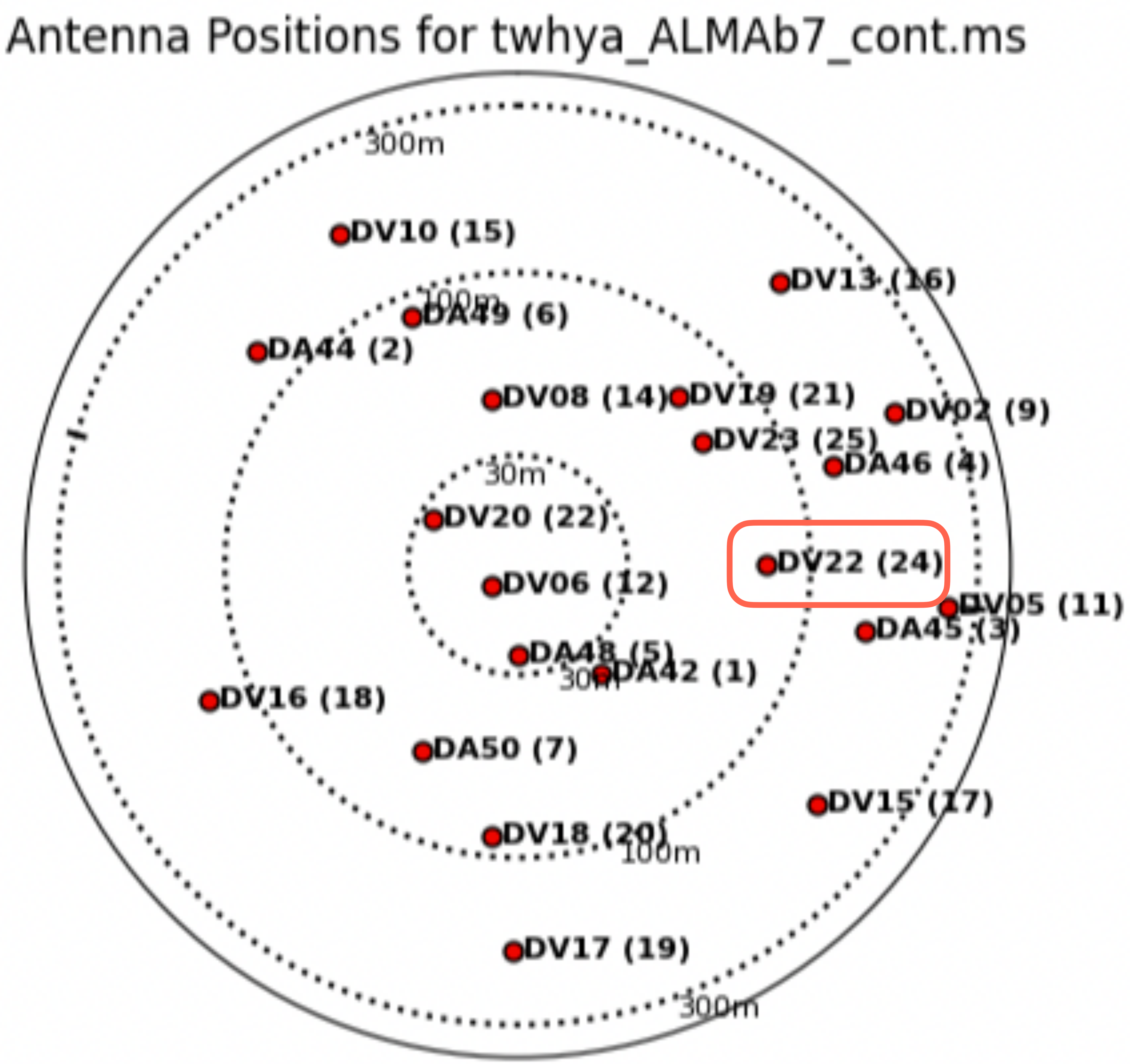
data	corrected	model
observations	self-cal data	CLEAN model
self-cal data		CLEAN model

used for
CLEAN imaging

“plotants” task in CASA

With “plotants” task in CASA:

```
plotants(vis =“twhya_ALMAb7_cont.ms",
              showgui = False,
              antindex = True,
              checkbaselines = True,
              logpos = True,
              figfile = “ant_positions.png")
```



“solint” in gaincal task

Solution Interval for `solint = 'inf'`

- The solution interval corresponds to the **longest continuous time span of data** within each scan.

In twhya_ALMAb7_cont.ms.listfile

Scan	Start Time (UTC)	End Time (UTC)	Duration (sec)
12	07:56:23.5	08:02:11.3	**347.8 sec**
16	08:08:09.6	08:13:57.3	**347.7 sec**
20	08:19:53.9	08:25:41.7	**347.8 sec**
24	08:32:00.5	08:37:48.2	**347.7 sec**
28	08:43:45.6	08:49:33.4	**347.8 sec**
36	09:05:15.6	09:07:31.6	**136.0 sec**

If you set **solint = 'inf'**, then the solution interval (time) will be the entire scan duration for each scan.
For scans 12, 16, 20, 24, and 28, the solution interval is 348 sec.
For scan 36, the solution interval is 136 sec.
Each scan will have a single gain solution over the respective time duration.

References

1. Cornwell et al. 1999
Self-calibration in Synthesis Imaging in Radio Astronomy II

2. Brogan et al. 2018
Advanced Gain Calibration Techniques in Radio Interferometry

3. Richards et al. 2022
ALMA Memo 620
Self-calibration and improving image fidelity for ALMA and other radio interferometers

4. First Look at Self Calibration CASA 6 from NRAO