EUROPEAN ARC
ALMA Regional Centre

ALMA Data

What comprises ALMA data and what's in your ALMA data

Luke Maud - ESO - Garching

credits also to - Carmen Toribio

SPANISH
ALMA Days

18-20 February 2025, La Laguna, Tenerife, Spain

Outline

- **INTRO : mm-interferometry**
  - aperture synthesis

what the observatory does *(QA talk next)*

- **PART 1 : What does ALMA data comprise of**
  - what sources are observed and why
  - top-level overview of calibration

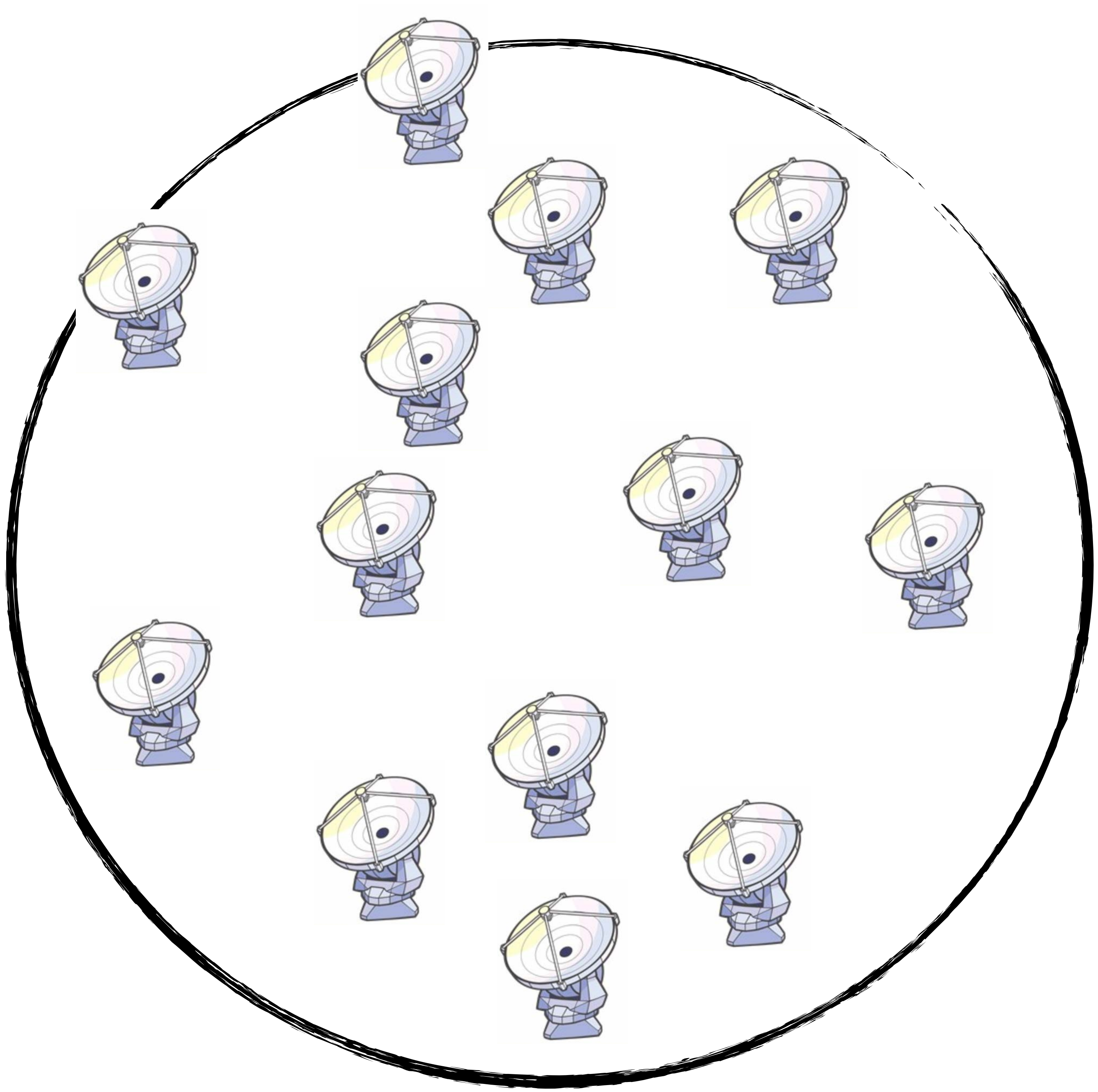
- **PART 2 : What is in your ALMA data**
  - what do you get
  - how is everything organised
  - how can you look at things

what you get to work with (hands-on session)

Aperture Synthesis

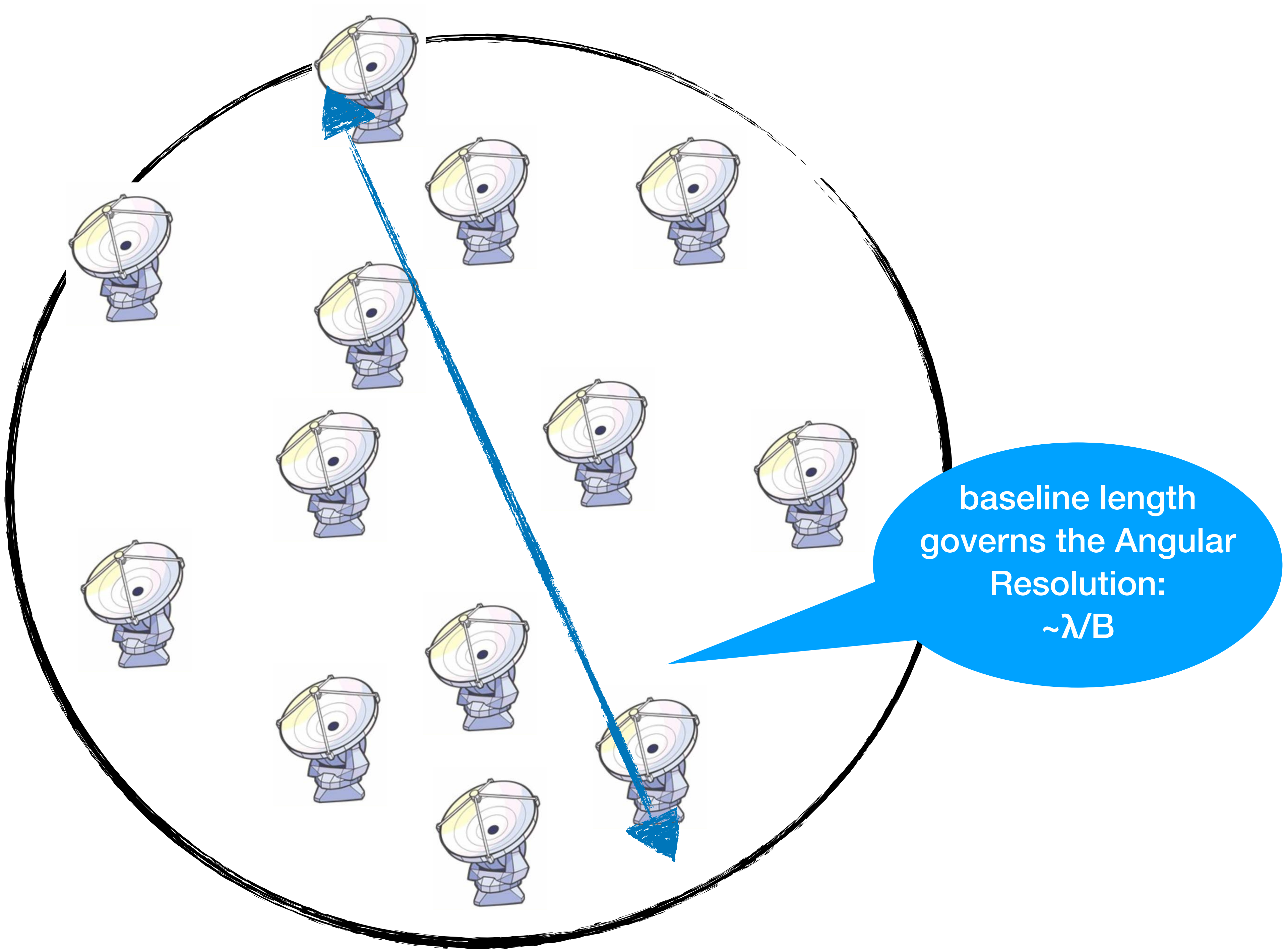
- combining multiple antennas to achieve an angular resolution equivalent to a very large filled single dish

- each baseline, or antenna pair measures a *single* spatial scale (size) - this is **one** component of a **Fourier** transform (power spectrum) of the 'on-sky' brightness distributions of an Astronomical Target



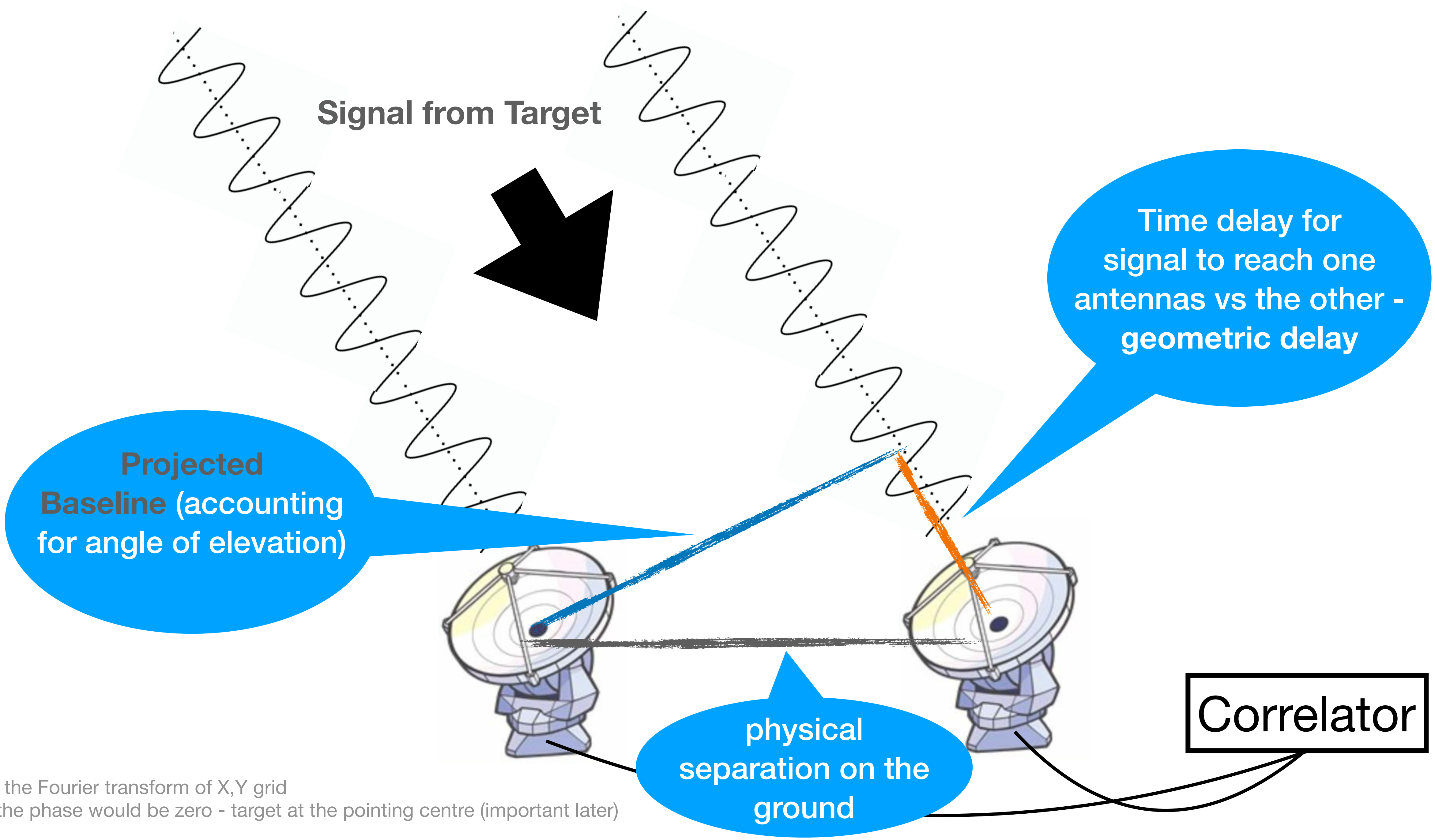
Aperture Synthesis

- combining multiple antennas to achieve an angular resolution equivalent to a very large filled single dish
- each baseline, or antenna pair measures a *single* spatial scale (size) - this is **one** component of a **Fourier** transform (power spectrum) of the 'on-sky' brightness distributions of an Astronomical Target



Aperture Synthesis

- each baseline records visibilities, a **complex number**, measured in **U,V*** space
- visibilities can be understood as an **amplitude** (flux) and **phase** (position**) component



*can be though of as the Fourier transform of X,Y grid
** for a point source the phase would be zero - target at the pointing centre (important later)

Aperture Synthesis

- each baseline records visibilities, a **complex number**, measured in **U,V*** space
- at any given time (during an observation) the baselines track over U,V space

Short snapshot (few min)

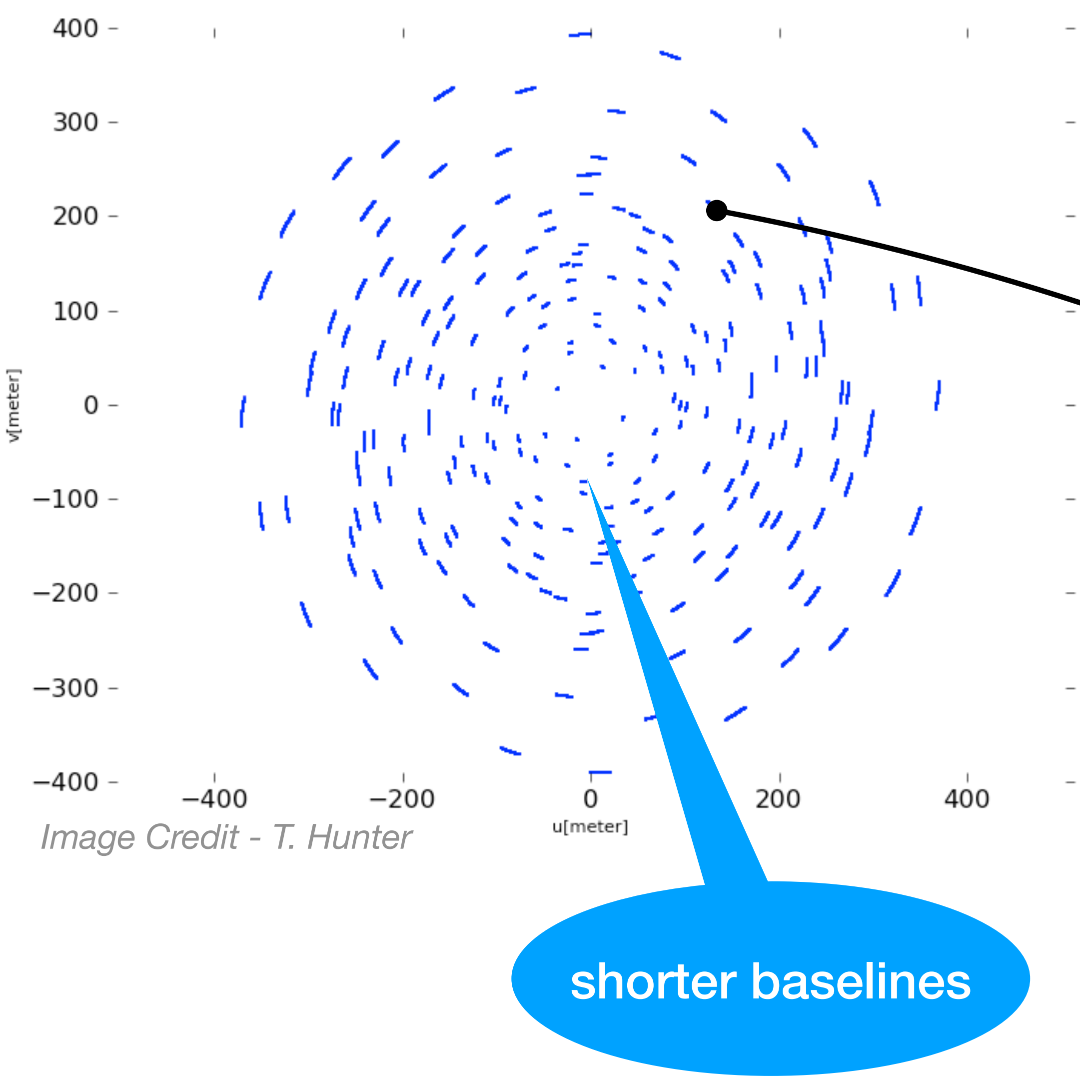
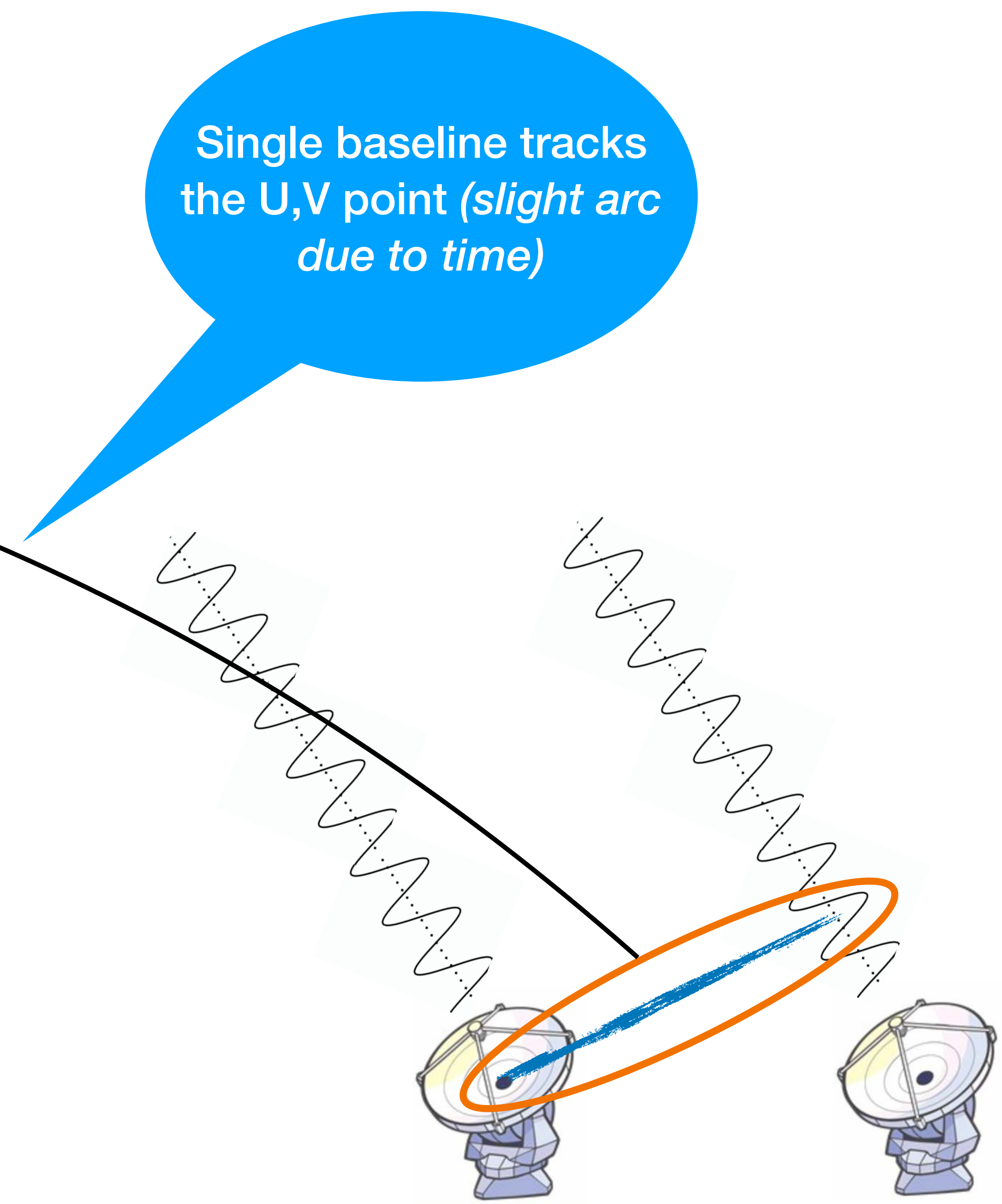


Image Credit - T. Hunter



*can be though of as the Fourier transform of X,Y grid

Aperture Synthesis

- each baseline records visibilities, a **complex number**, measured in **U,V*** space
- at any given time (during an observation) the baselines track over U,V space

Short snapshot (few min)

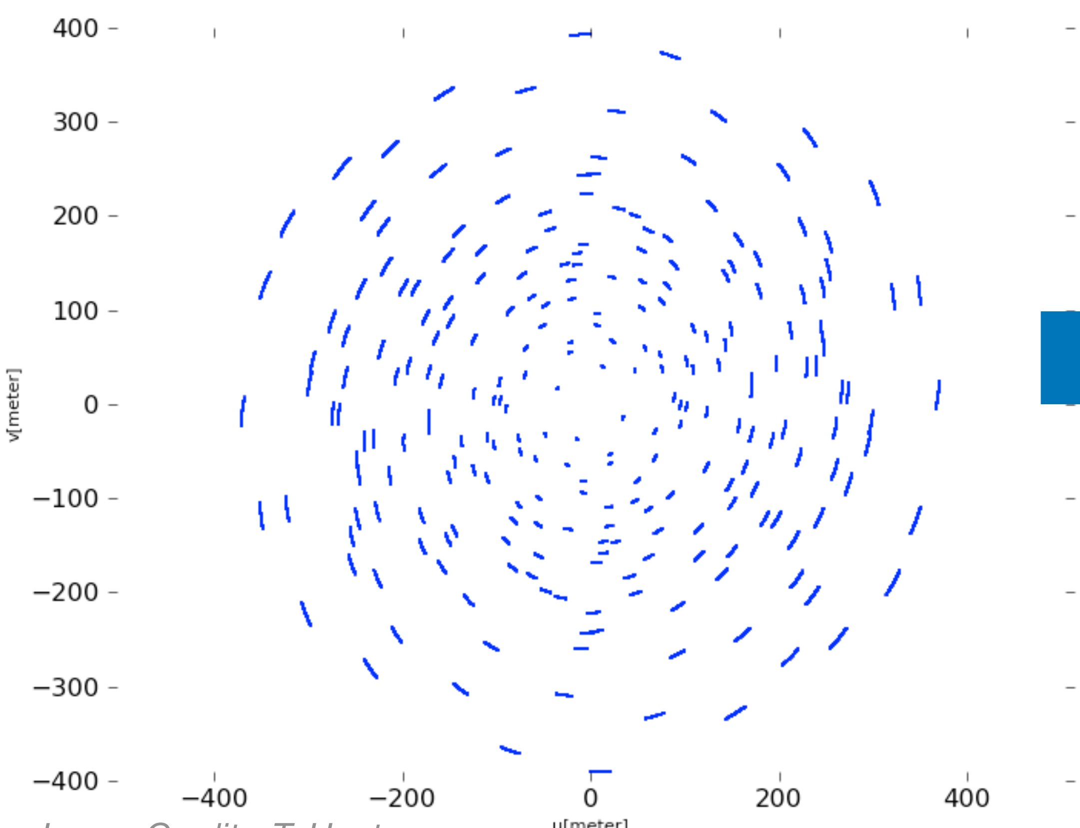
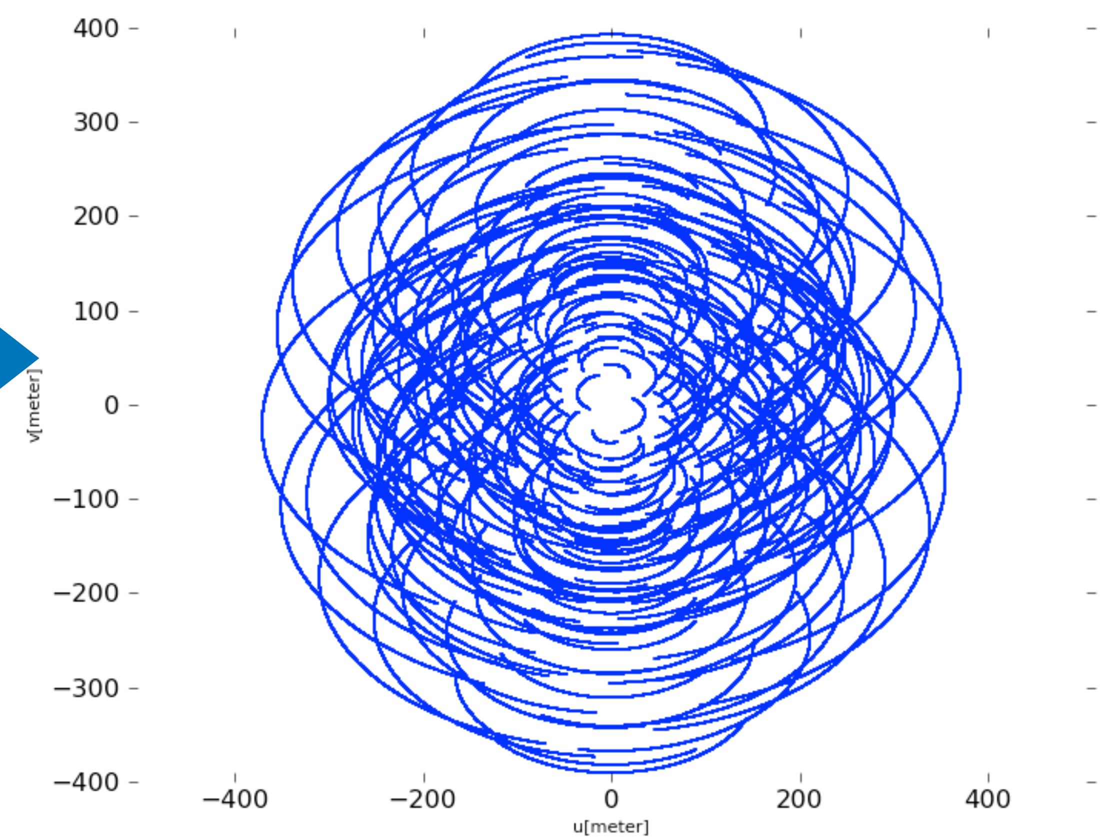


Image Credit - T. Hunter

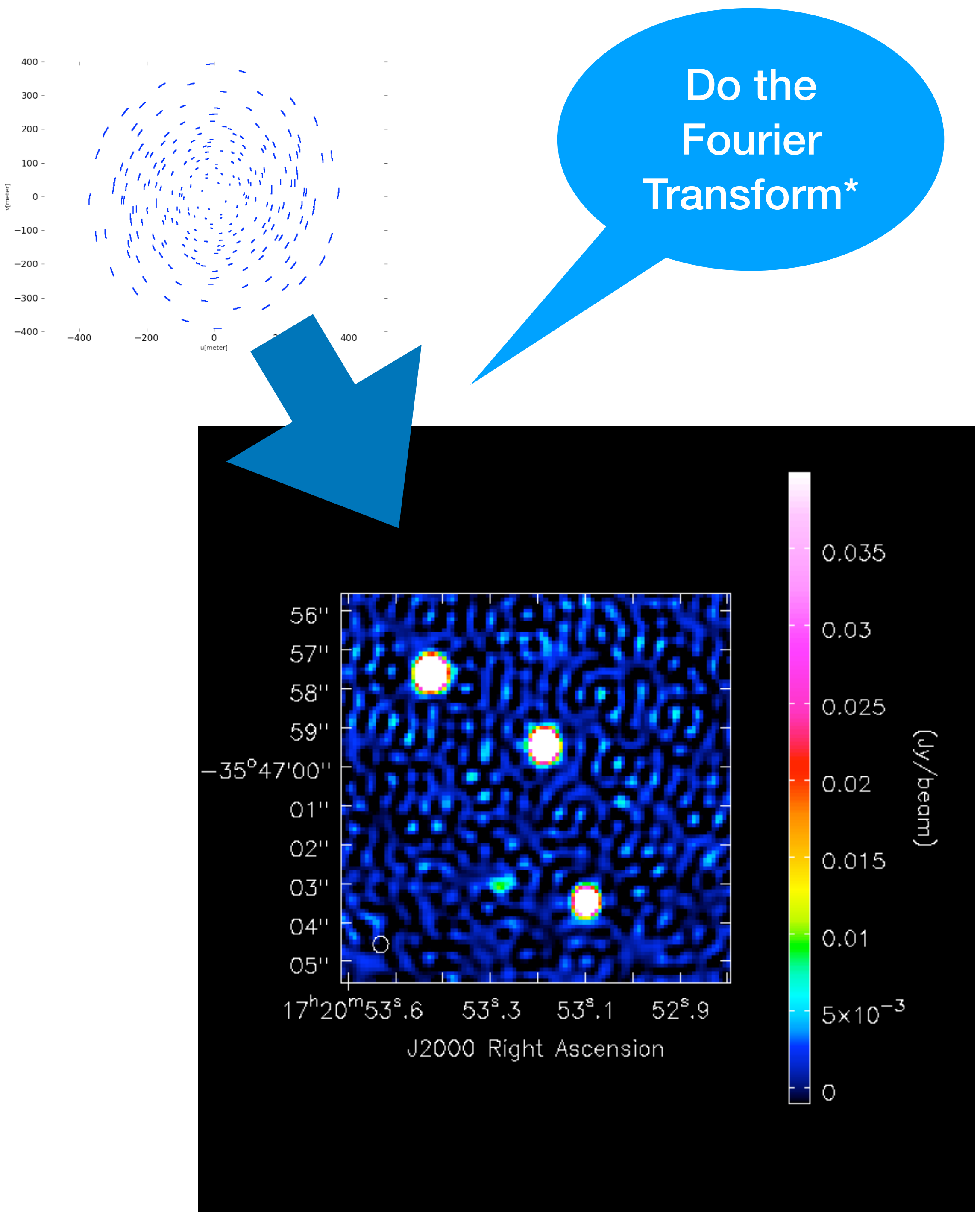
Let the Earth rotate - ~2hours

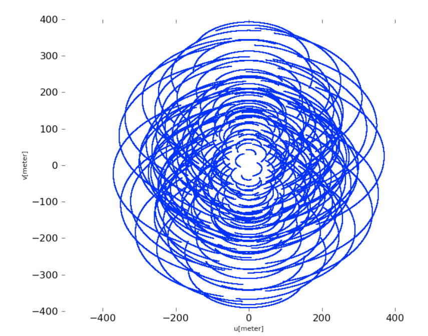


*can be though of as the Fourier transform of X,Y grid

Aperture Synthesis

- a better **filled** U,V coverage will better represent the true on sky brightness distribution



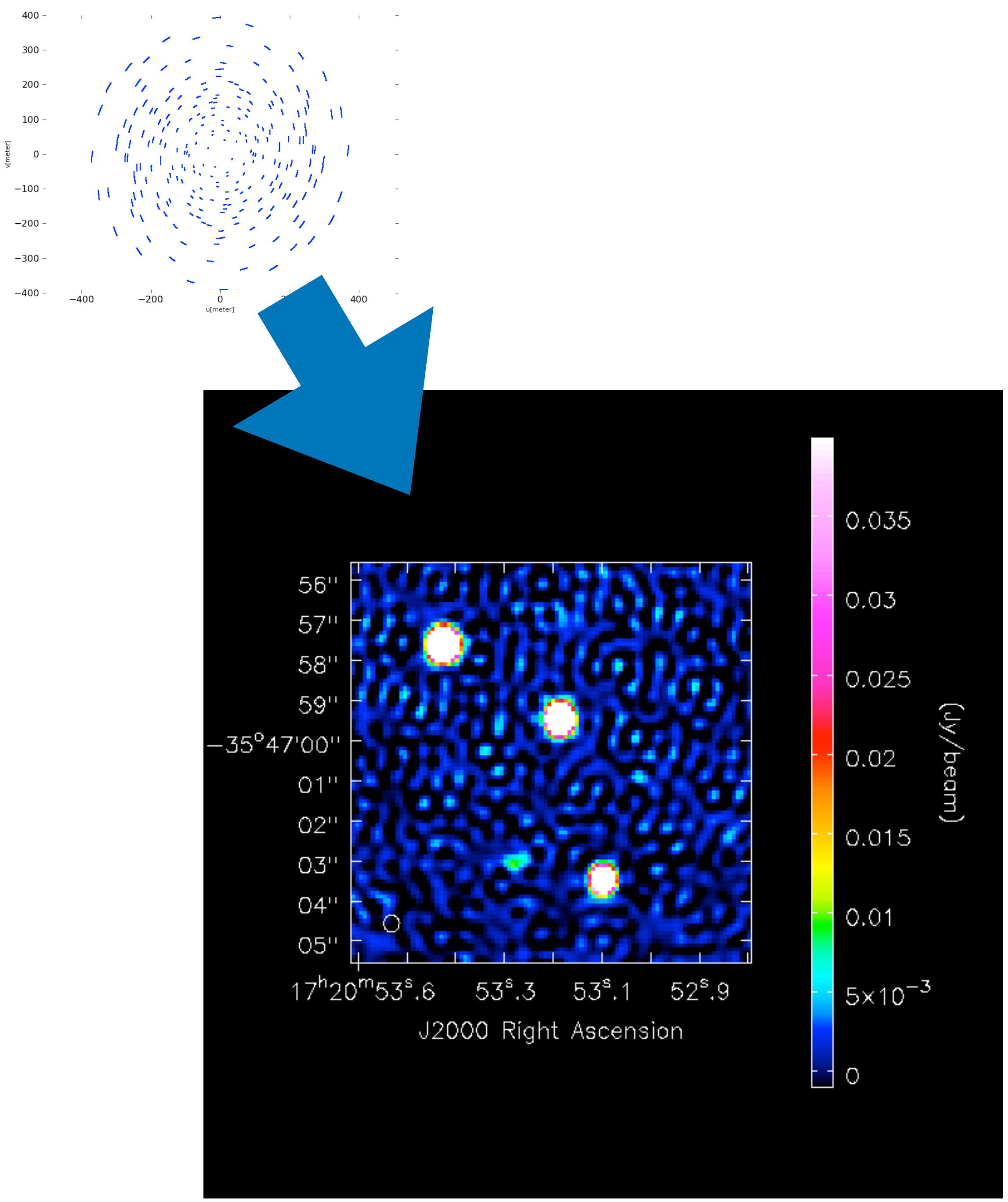


*van Cittert-Zernike theorem

Image Credit - T. Hunter

Aperture Synthesis

- a better **filled** U,V coverage will better represent the true on sky brightness distribution



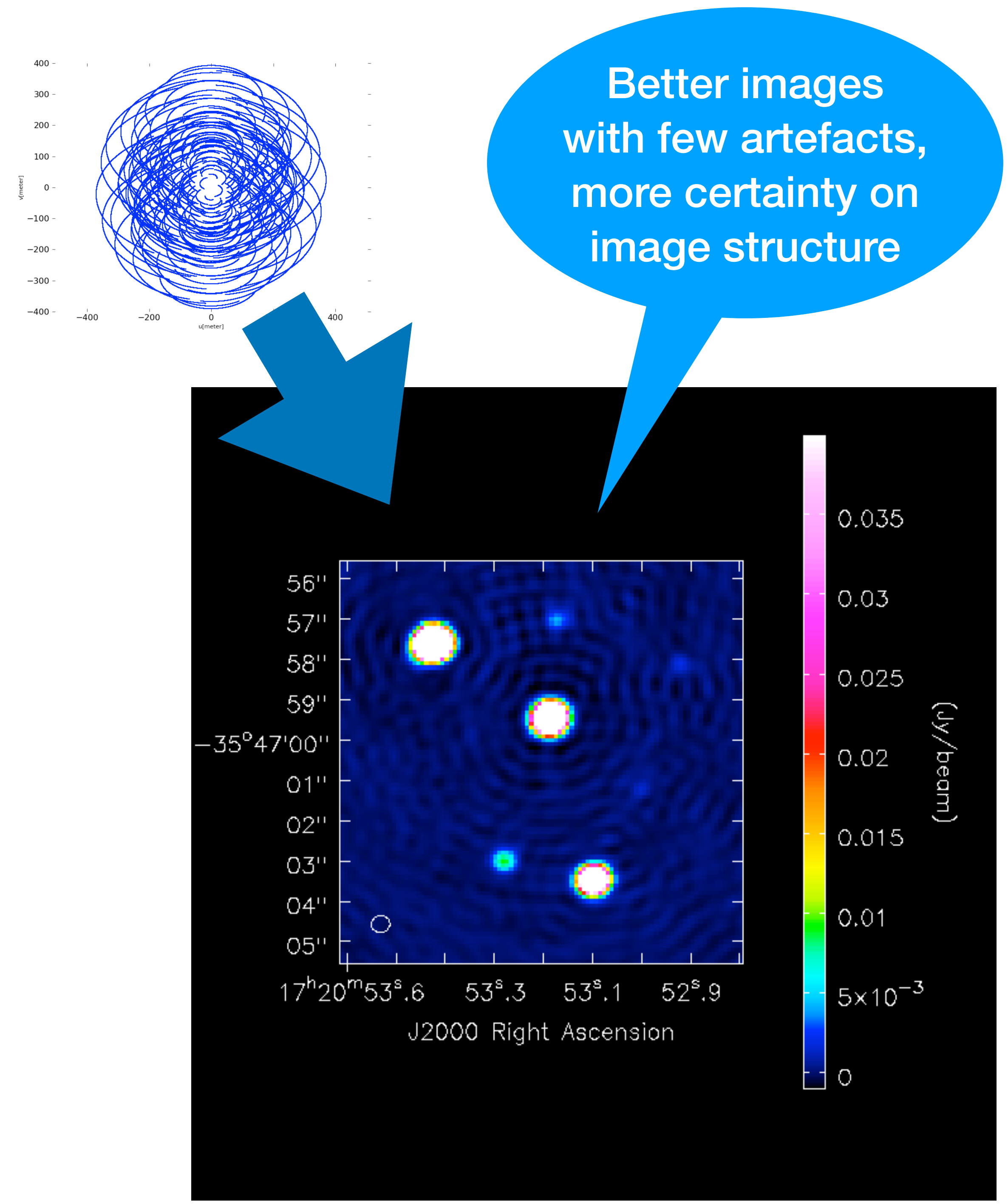
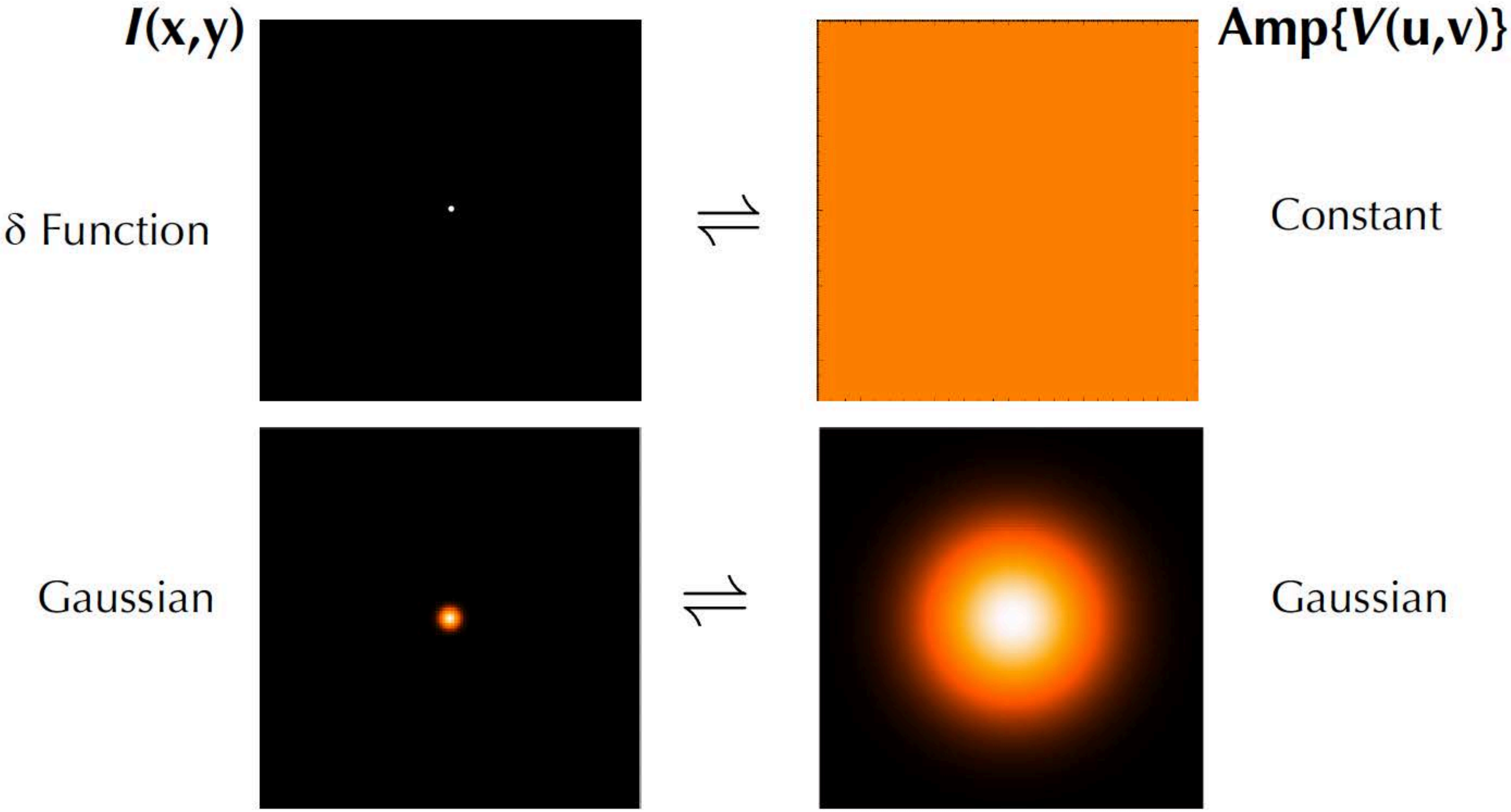


Image Credit - T. Hunter

Some 2-D Fourier Transform Pairs



narrow features transform to wide features (and vice-versa)

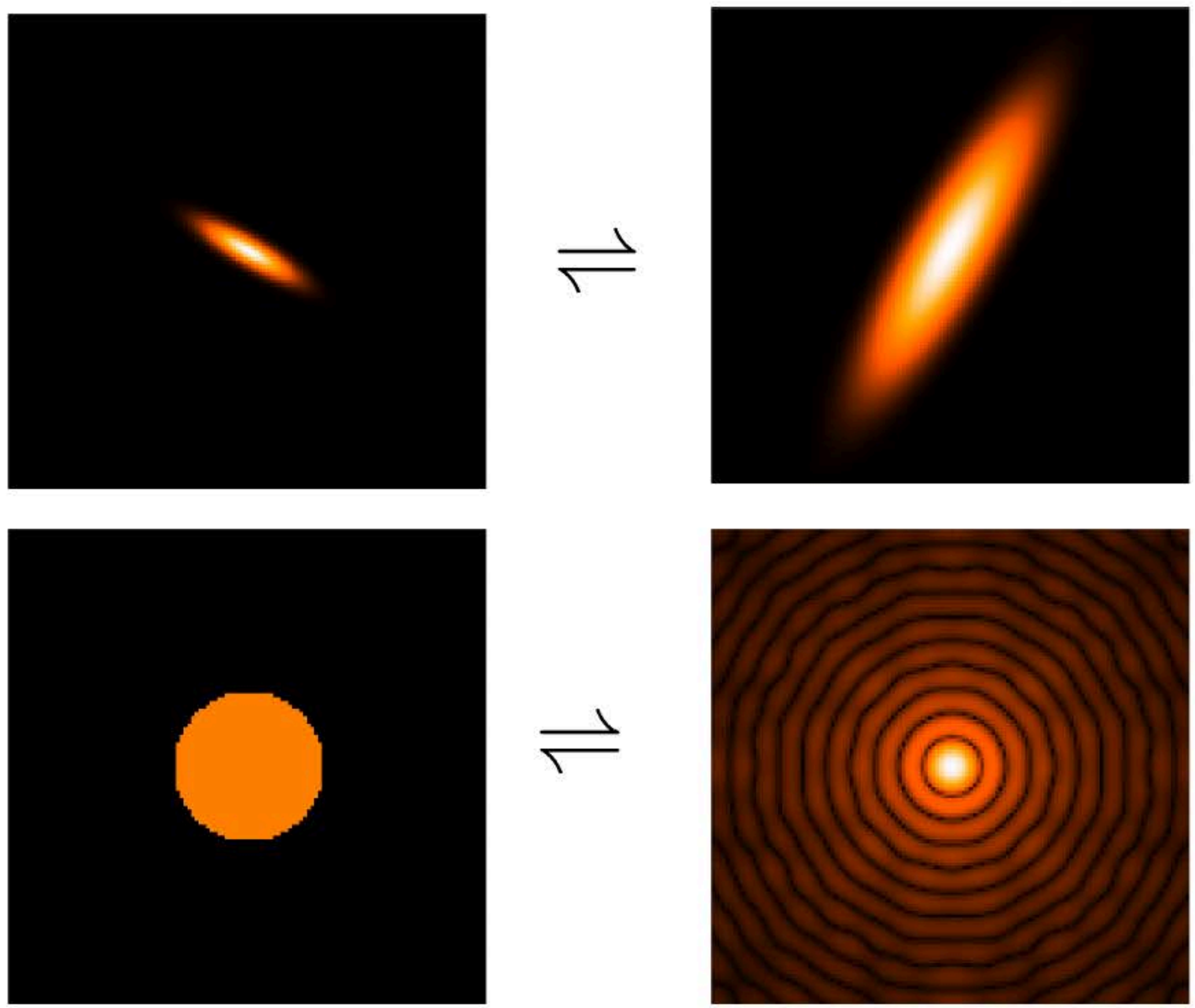
J. Wilner

More 2-D Fourier Transform Pairs

$I(x,y)$ | Amp{$V(u,v)$}

elliptical Gaussian ⇌ elliptical Gaussian

Disk ⇌ Bessel



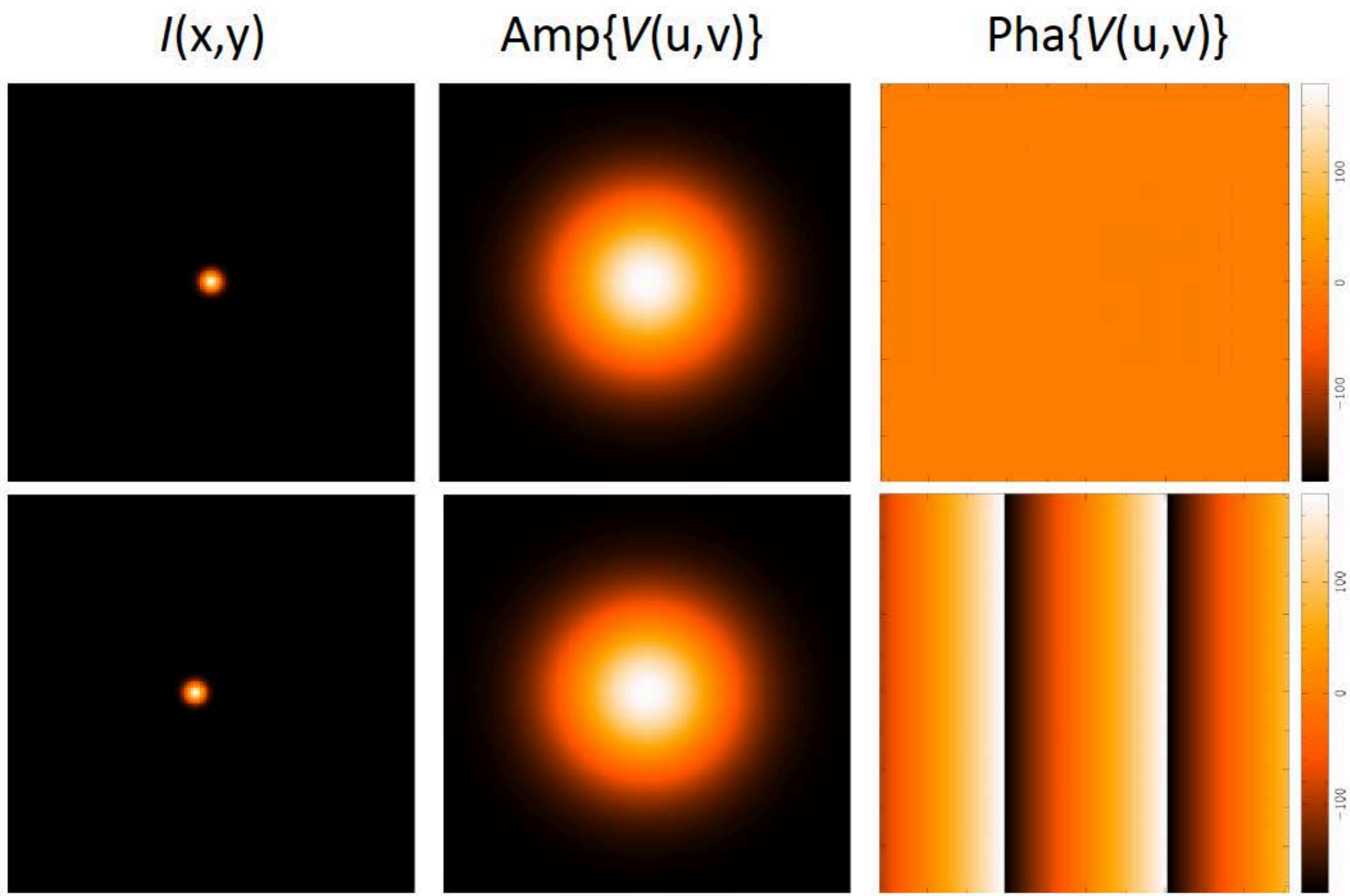
sharp edges result in many high spatial frequencies

J. Wilner

Amplitude and Phase

Complex numbers: (real, imaginary) or (amplitude, phase)

- amplitude tells “how much” of a certain spatial frequency component
- phase tells “where” this component is located



J. Wilner

Any questions at this stage?

- **PART 1 : What does ALMA data comprise of**
  - what sources are observed and why
  - basic overview of calibration

Part 1

ALMA Technical Handbook

an overview of all material ALMA related is here

Doc 11.3, version 1.4 | March 1st, 2024

ALMA Cycle 11 Technical Handbook

ALMA

www.almascience.org

ing its member states), NSF (USA) and NINS (Japan), together with NRC
d KASI (Republic of Korea), in cooperation with the Republic of Chile.

more detail about data and calibration than I can cover here

Chapter 10

Calibration and Calibration Strategies

This chapter describes the methods and philosophy used by ALMA in order to calibrate the correlated visibility function data. Since calibration provides a central and important part of ALMA's production of images, this chapter contains many references to other chapters that describe other aspects of the ALMA systems. In particular, the reader of this chapter will want to be familiar with the Principles and Concepts of Interferometry described in Chapter 3 and the ALMA Observing Modes described in Chapter 8. It may also be useful to be familiar with the Quality Assurance process described in Chapter 11.

10.1 Fundamental Synthesis Relationship

The relationships between the visibility function $\mathcal{V}(u,v)$ and the sky emission $I(l,m)$, embodied in the van Cittert-Zernike theorem (see Chapters 3 and 7), are:

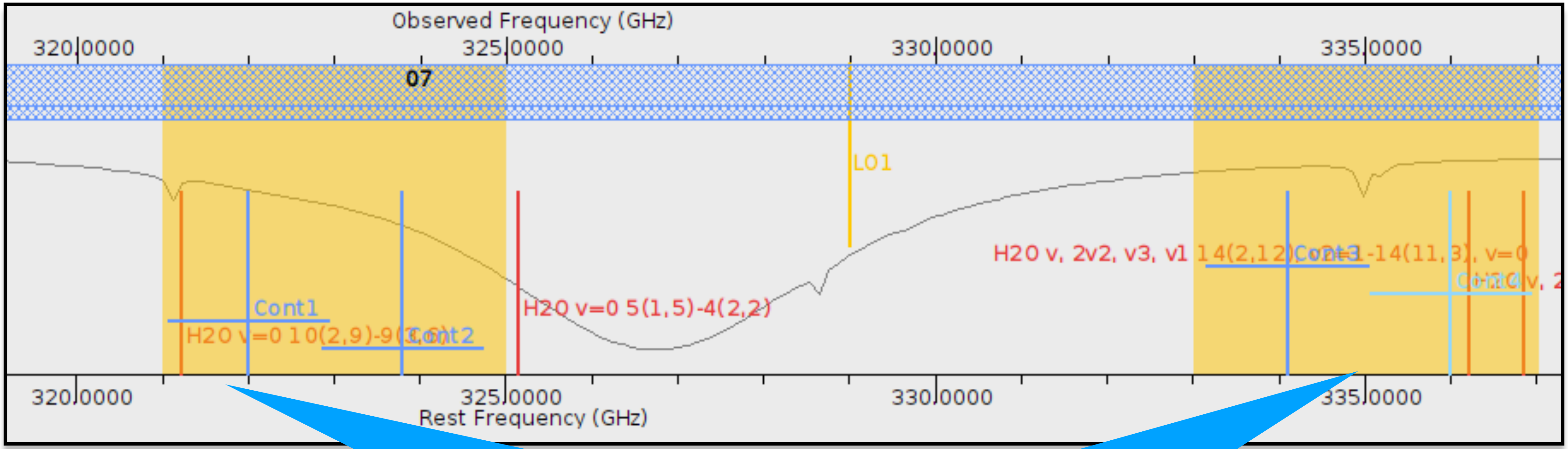
$$\mathcal{V}(u,v) = \iint A(l,m)I(l,m)e^{2\pi i(ul+vm)}\,dldm = Ae^{i\phi} \tag{10.1}$$

$$A(l,m)I(l,m) = \iint \mathcal{V}(u,v)e^{-2\pi i(ul+vm)}\,dudv \tag{10.2}$$

The visibility function, $\mathcal{V}$, is thus the summation of the emission distribution of the source $I(l,m)$ (where (l,m) are its direction cosines), convolved by the exponential term that is the delay difference of the signal from the source to each antenna. The primary beam term, $A(l,m)$, describes the relative sensitivity of the antennas and is only a few arcminutes in size. The (u,v) *spatial coordinates* are not the physical separation of the two

What sources are observed and why

- **Target : We want to do science**
  - of course we need to look at our target of interest
  - data would be recorded in our selected Spectral Windows

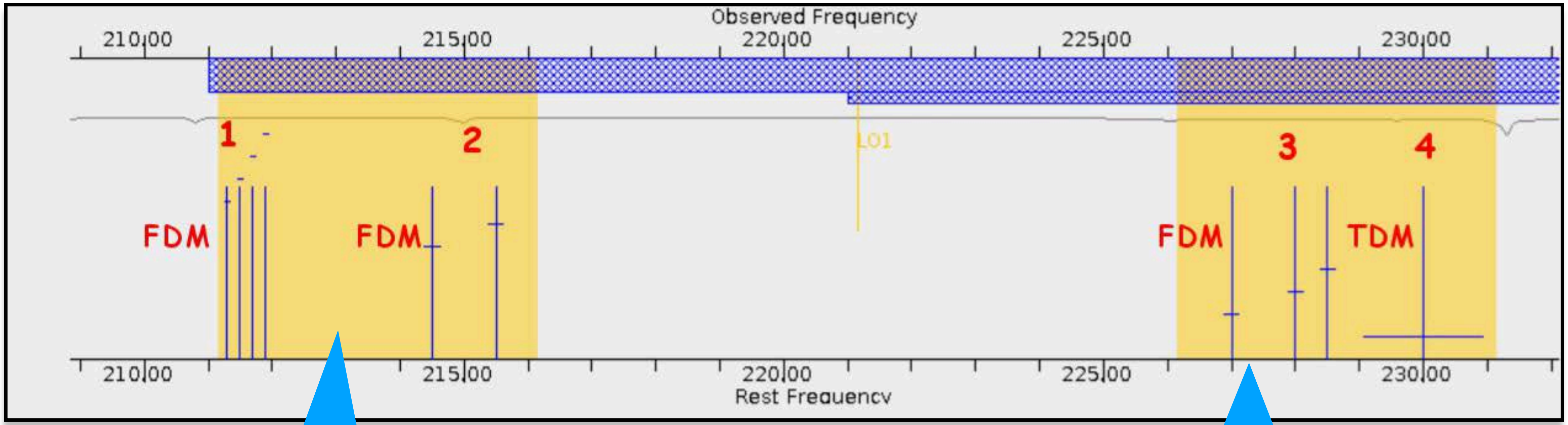


simple
setup with 4x wide ~2GHz
SpWs trying to cover H_2O

Part 1

What sources are observed and why

- **Target : We want to do science**
  - of course we need to look at our target of interest
  - data would be recorded in our selected Spectral Windows



complex setup with two basebands splitting into 4x and 2x groups of narrow SpWs

complex setup with one baseband splitting into 3x and one as full wide SpW

Part 1

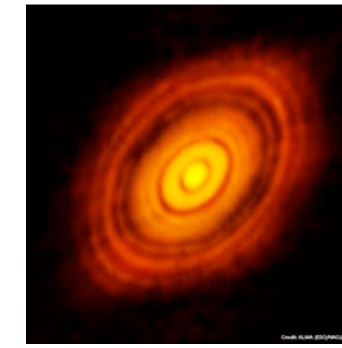
What sources are observed and why

- **Target : We want to do science**
  - of course we need to look at our target of interest
  - data would be recorded in our selected Spectral Windows

and……we use many antennas, what about instrumental effects or the atmosphere?

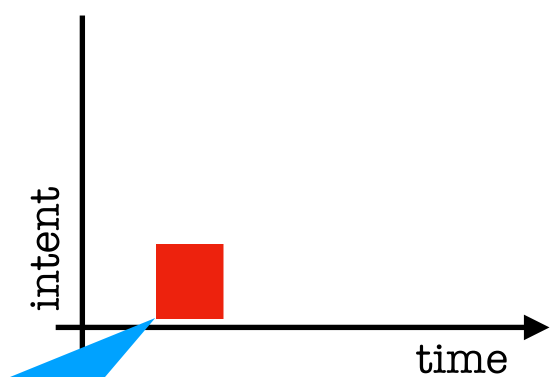
Part 1

What sources are observed and why



- **Target : We want to do science**
  - of course we need to look at our target of interest
  - data would be recorded in our selected Spectral Windows

- **Bandpass : We measure over a frequency range**
  - very bright **point-source*** which has no spectral features over our range of interest
  - corrects variations in antennas and signal path (receivers etc)



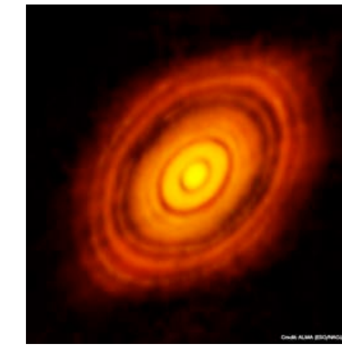
so-called "scan" of a particular intent. Made up of many short 3-6s "integrations"

*we know how the visibility PHASE should look

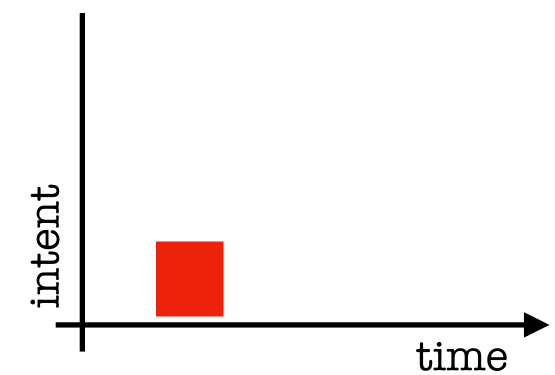
Part 1

What sources are observed and why

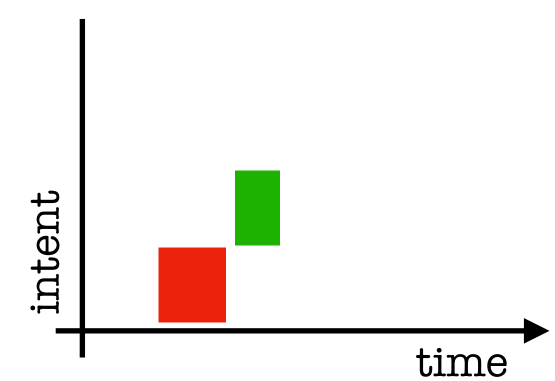
- **Target : We want to do science**
  - of course we need to look at our target of interest
  - data would be recorded in our selected Spectral Windows



- **Bandpass : We measure over a frequency range**
  - very bright **point-source*** which has no spectral features over our range of interest
  - corrects variations in antennas and signal path (receivers etc)



- **Flux : We need to have a corrected flux scale**
  - observe a known ‘flux’ point-like or solar system amplitude calibrator**



*we know how the visibility PHASE should look

**can often use the Bandpass if we know it’s flux well

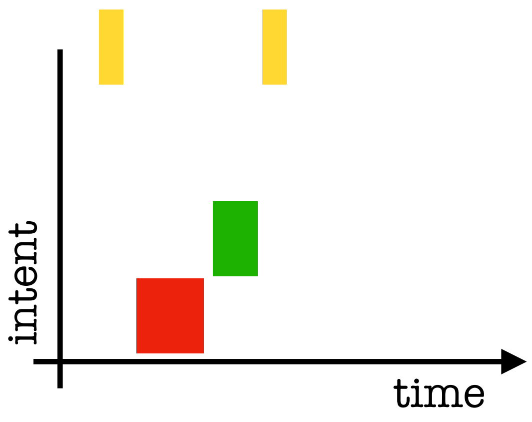
Part 1

What sources are observed and why

- **Pointing : We have to ensure we look in the correct direction**

  - the telescope is calibrated to know positions, but we have to check during the observations

  - if antennas don't respond correctly we need to know



Part 1

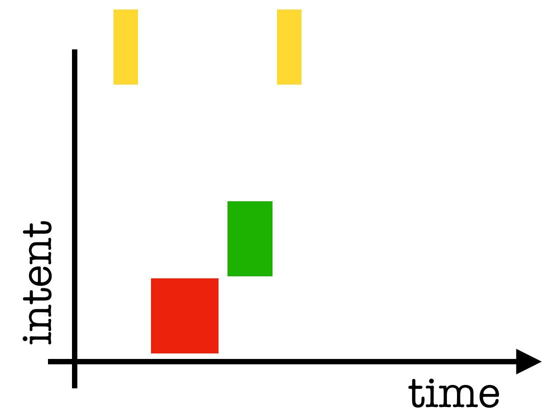
What sources are observed and why

- **Pointing : We have to** ... **rect direction**

  - the telescope is calib ... we have to check during the

  - if antennas don't respond cor ... y we need to know

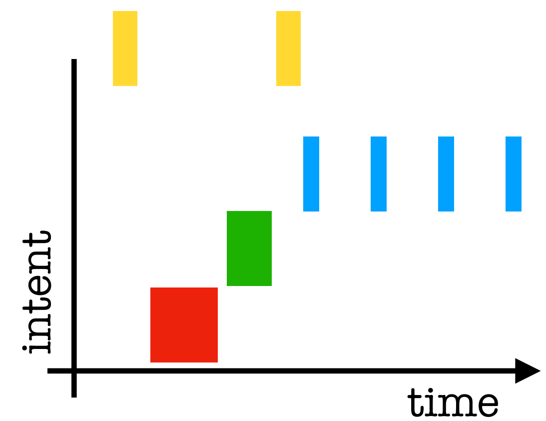
oh yes, this is also fundamental !



- **Gain Calibrator : We are looking through the atmosphere**

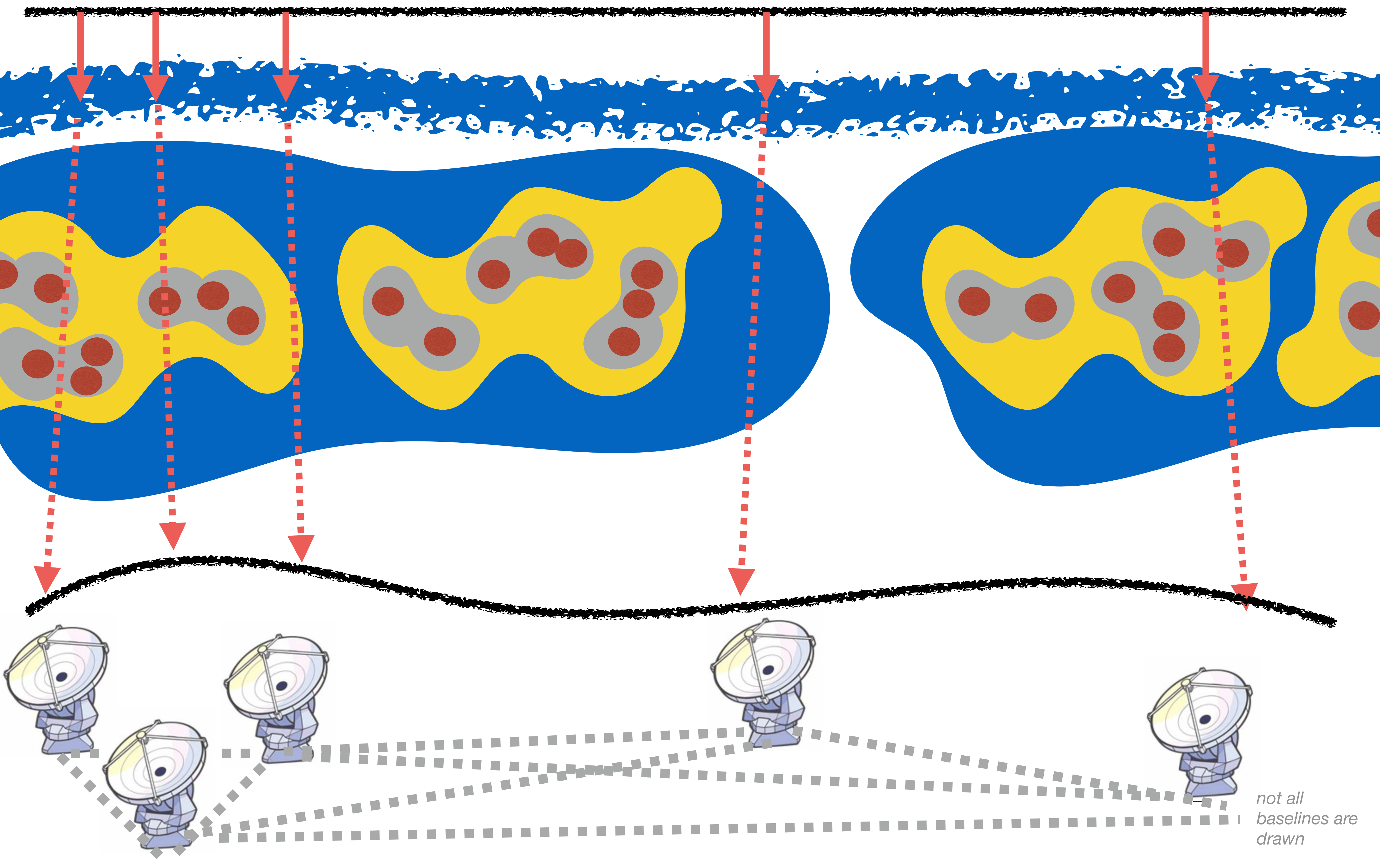
  - changes in amplitude and refraction caused by the troposphere with time

  - look regularly at a **point-source** calibrator



Part 1
Troposphere

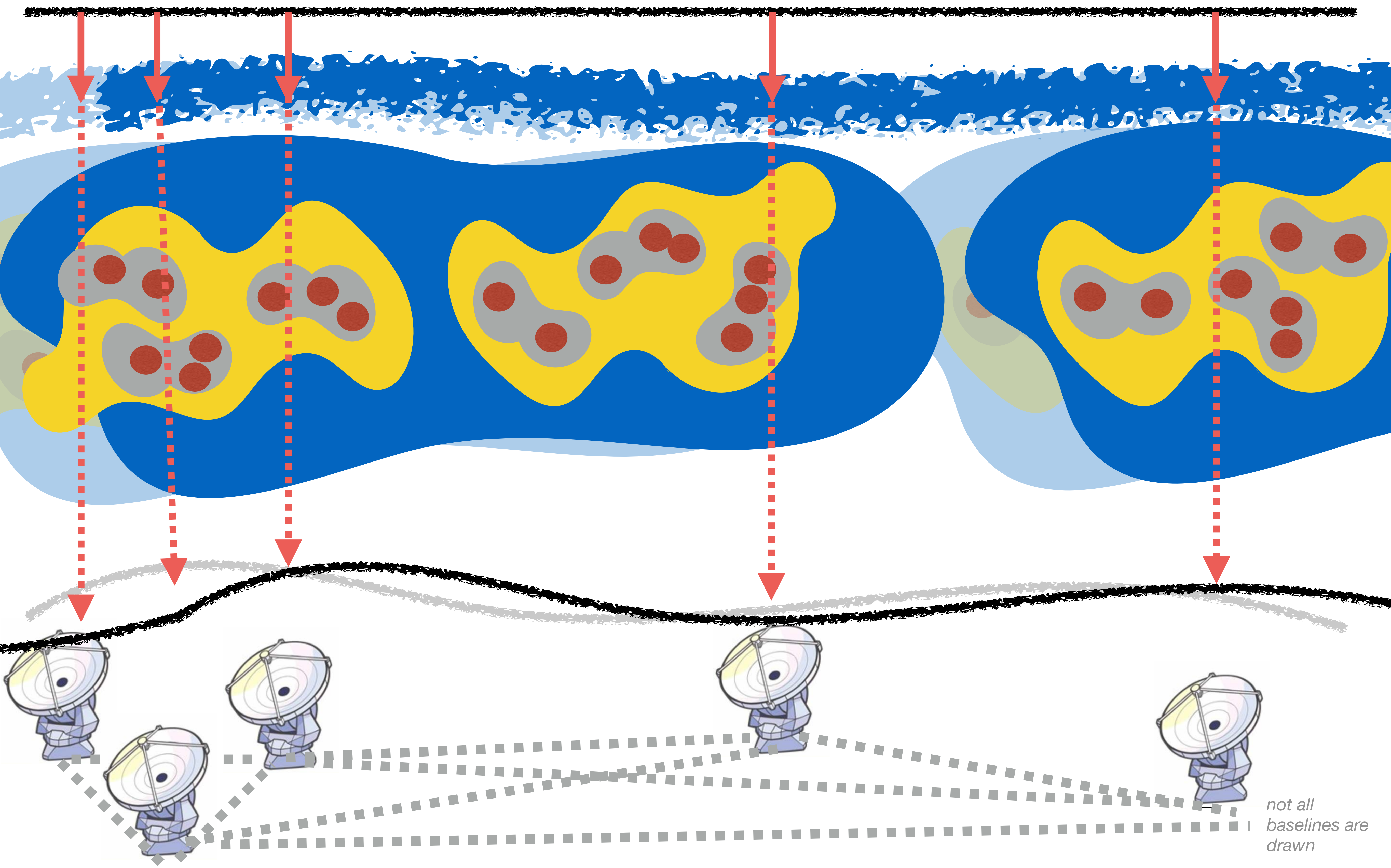
different baselines see different fluctuations causes by variable atmospheric ‘cells’, changes the arrival of the wavefront



Part 1

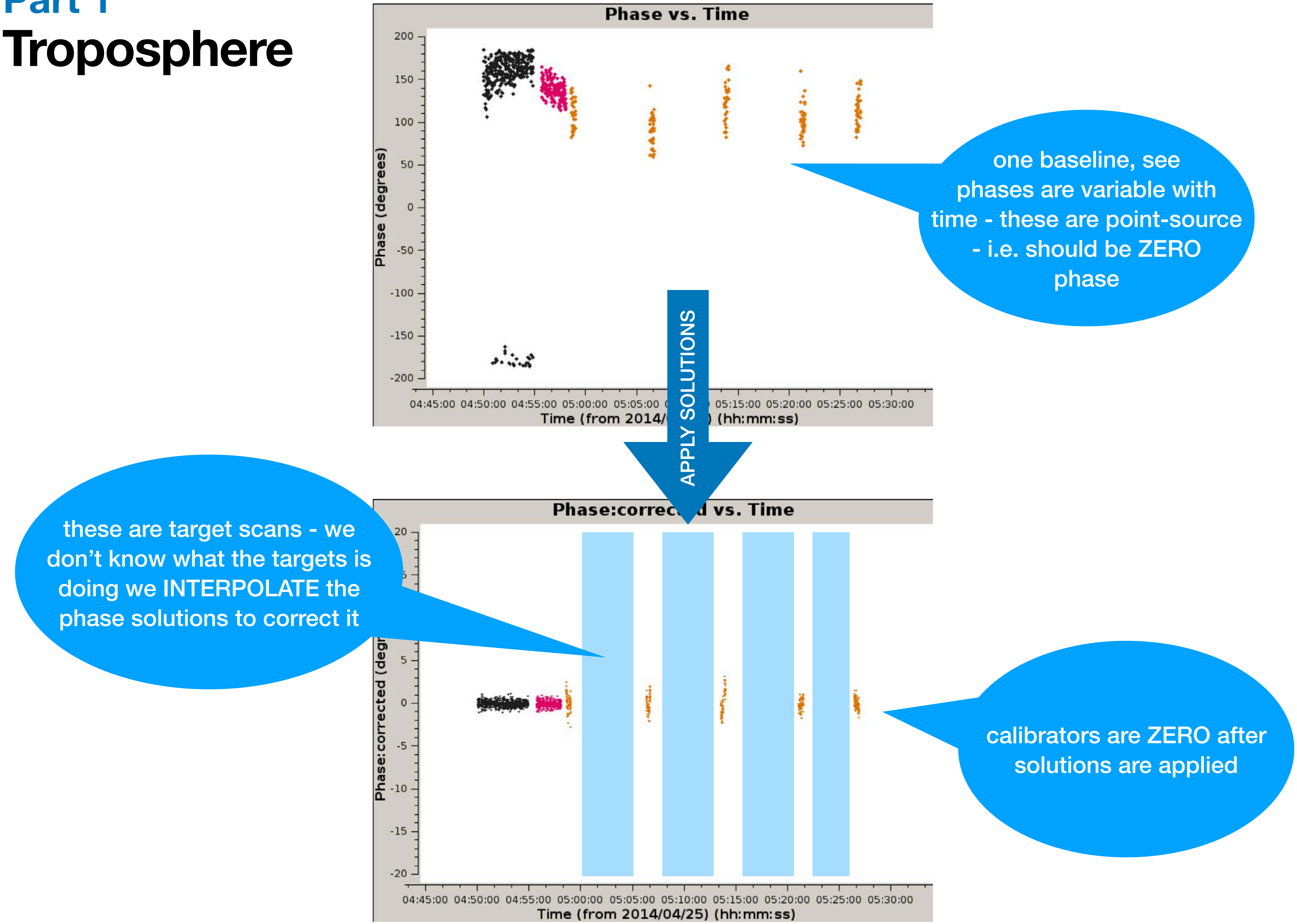
Troposphere

the tropospheric 'layer' moves with time (wind) ➡ variable wavefront ➡ variable **PHASE** changes



Part 1

Troposphere



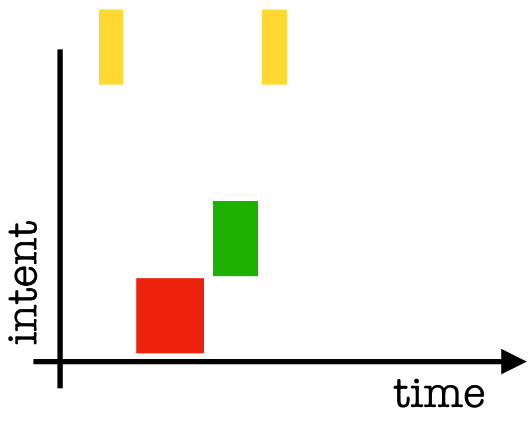
Part 1

What sources are observed and why

- **Pointing : We have to ensure we look in the correct direction**

  - the telescope is calibrated to know positions, but we have to check during the observations

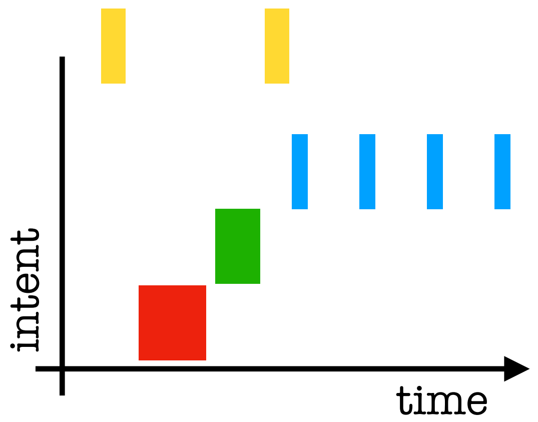
  - if antennas don't respond correctly we need to know



- **Gain Calibrator : We are looking through the atmosphere**

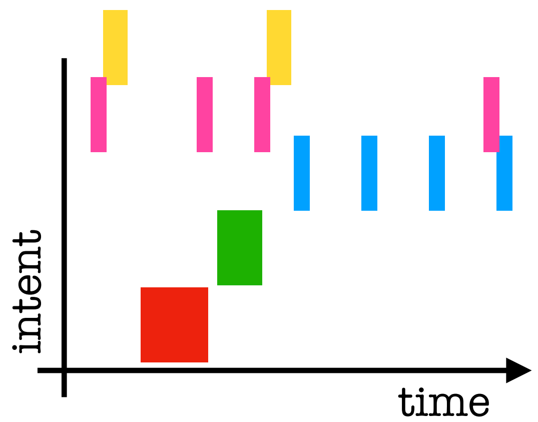
  - changes in amplitude and refraction caused by the troposphere with time

  - look regularly at a **point-source** calibrator



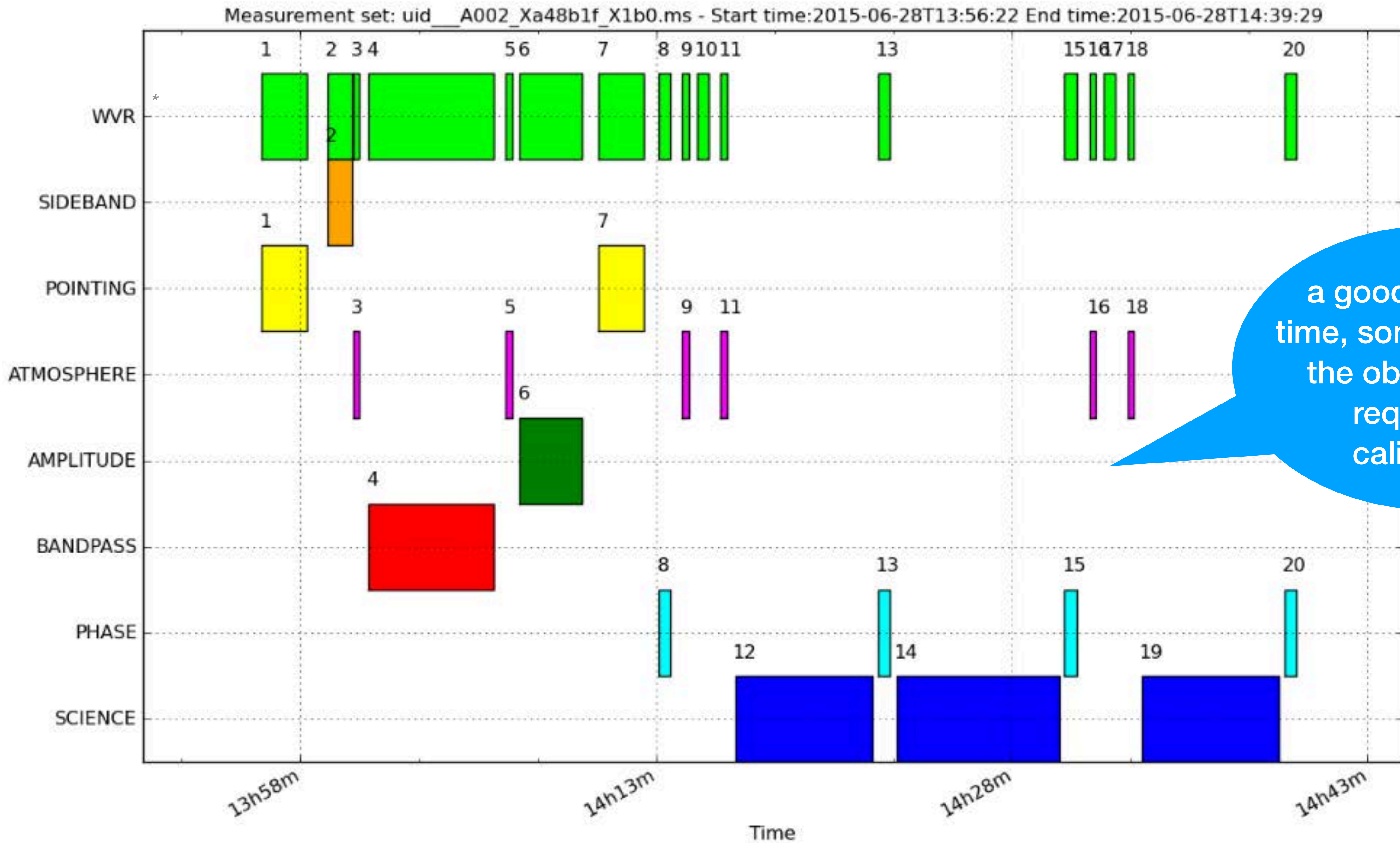
- **System Temperature : Amplitude scale correction for receiver and sky**

  - correctly scale from instrument units to flux also accounting for looking through the atmosphere



Part 1

What sources are observed and why

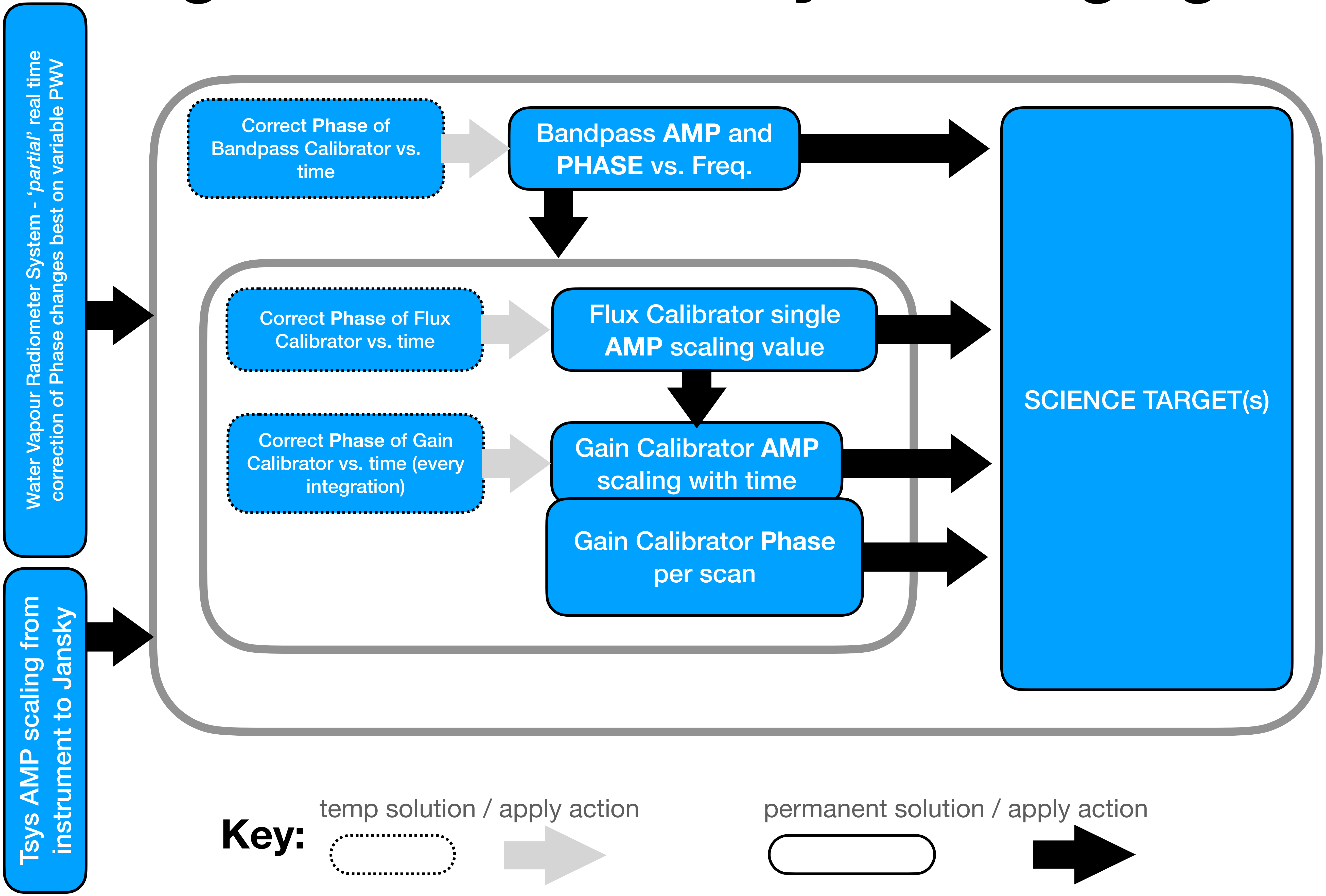


a good fraction of time, sometimes ~half the observation is required for calibrations

** ALMA also can measure the PWV content on different antennas and also proves a 'phase' corrected - this is the WVR system -this is always 'ON' for the 12m array*

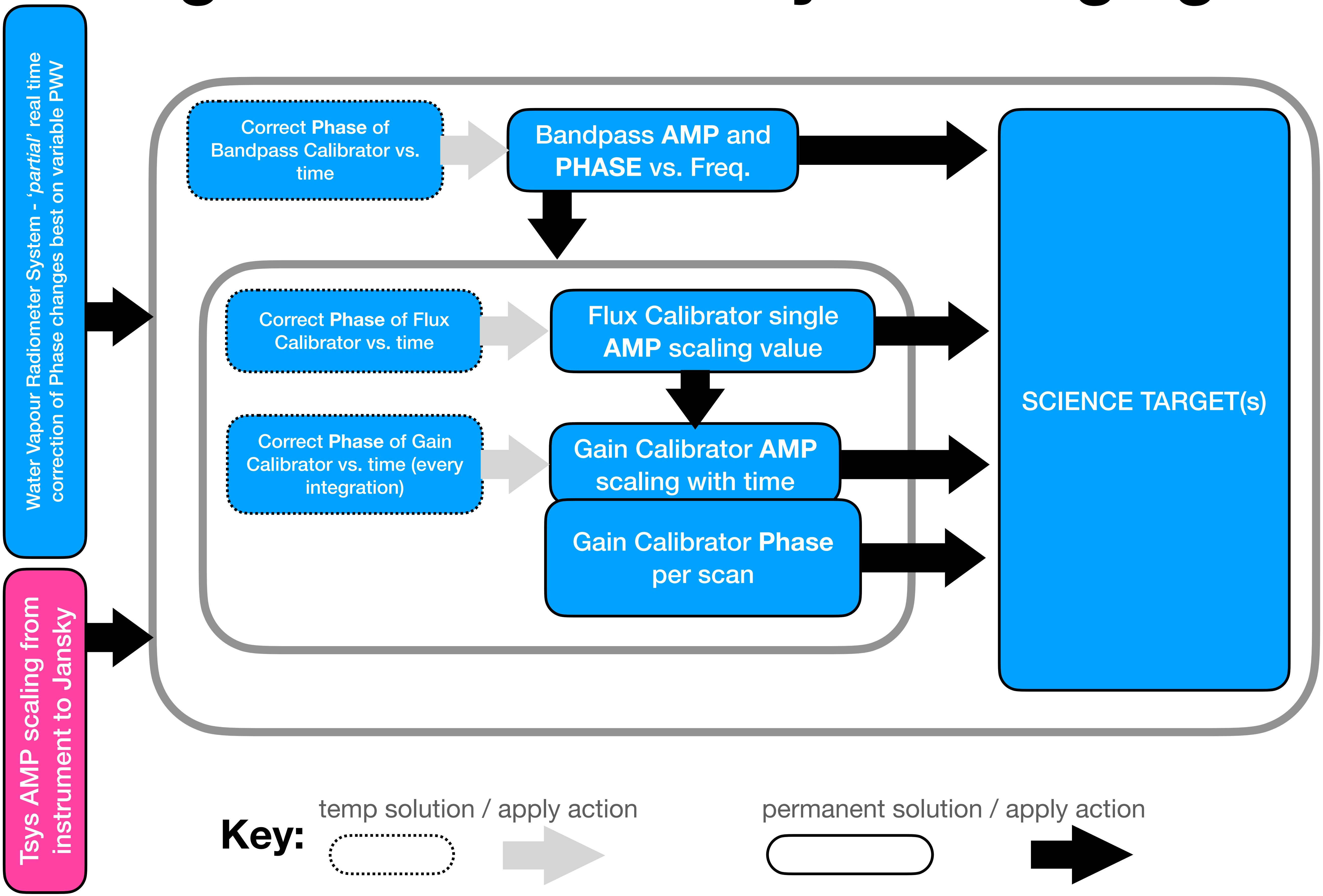
Part 1

Data get calibrated ready for imaging



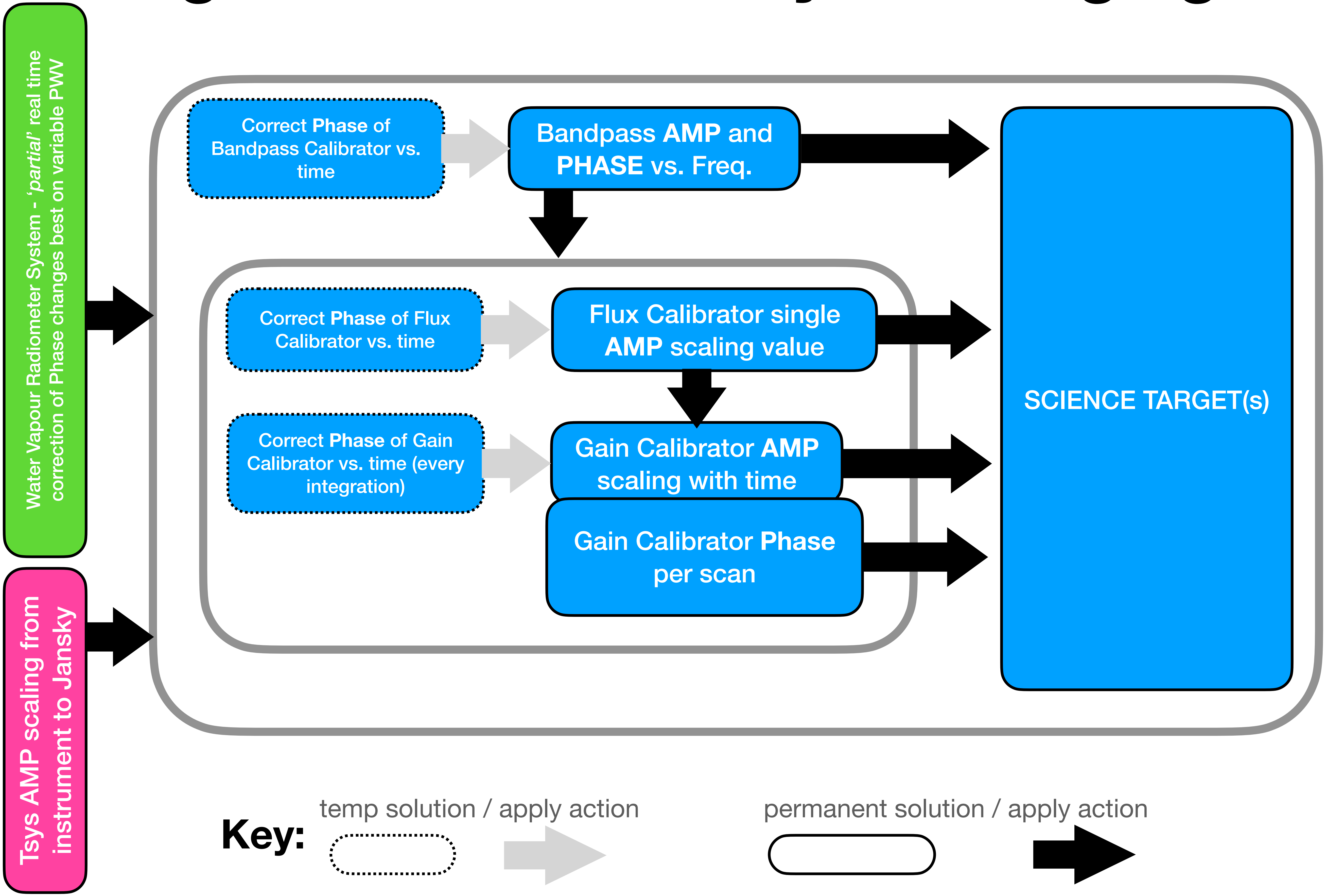
Part 1

Data get calibrated ready for imaging



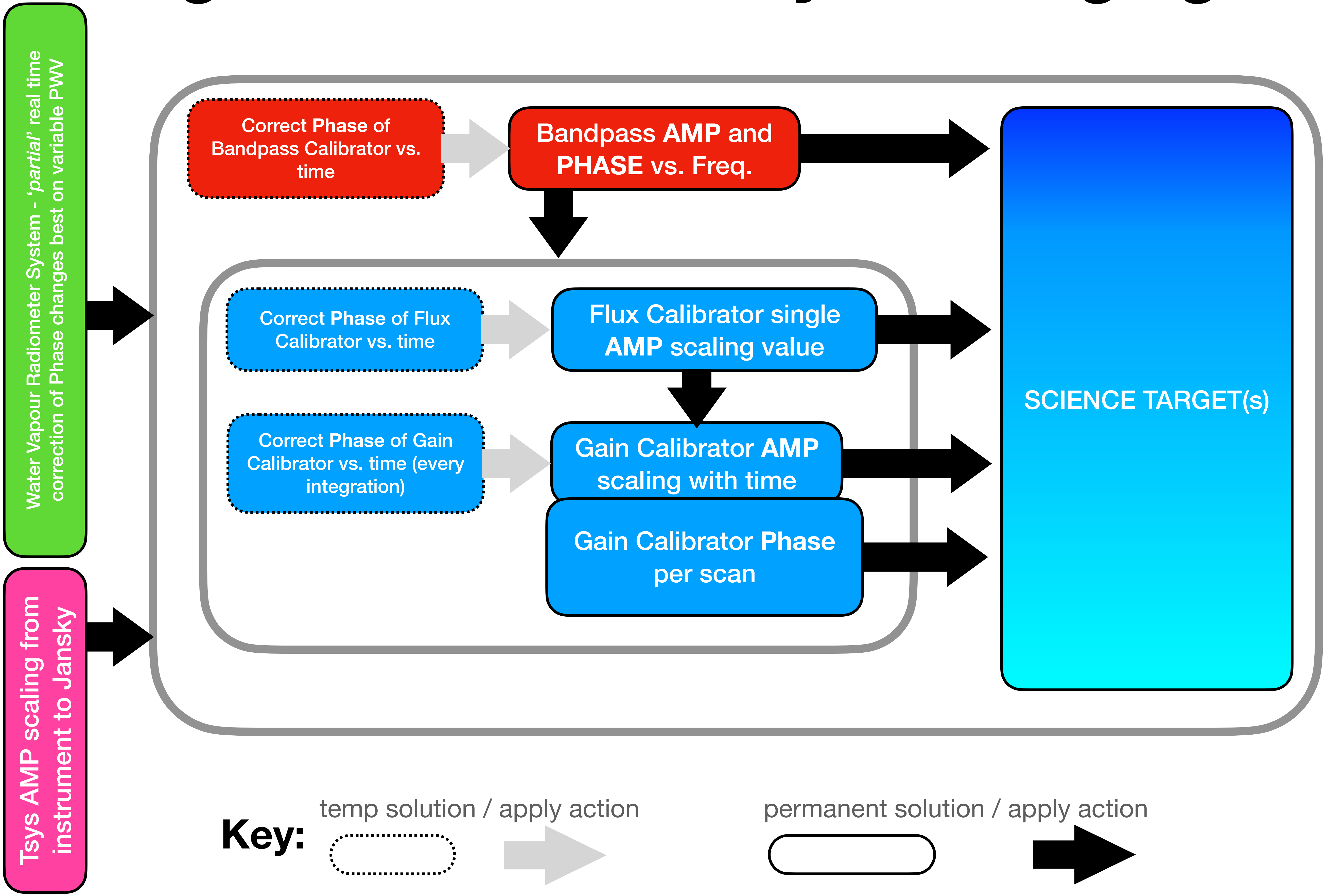
Part 1

Data get calibrated ready for imaging



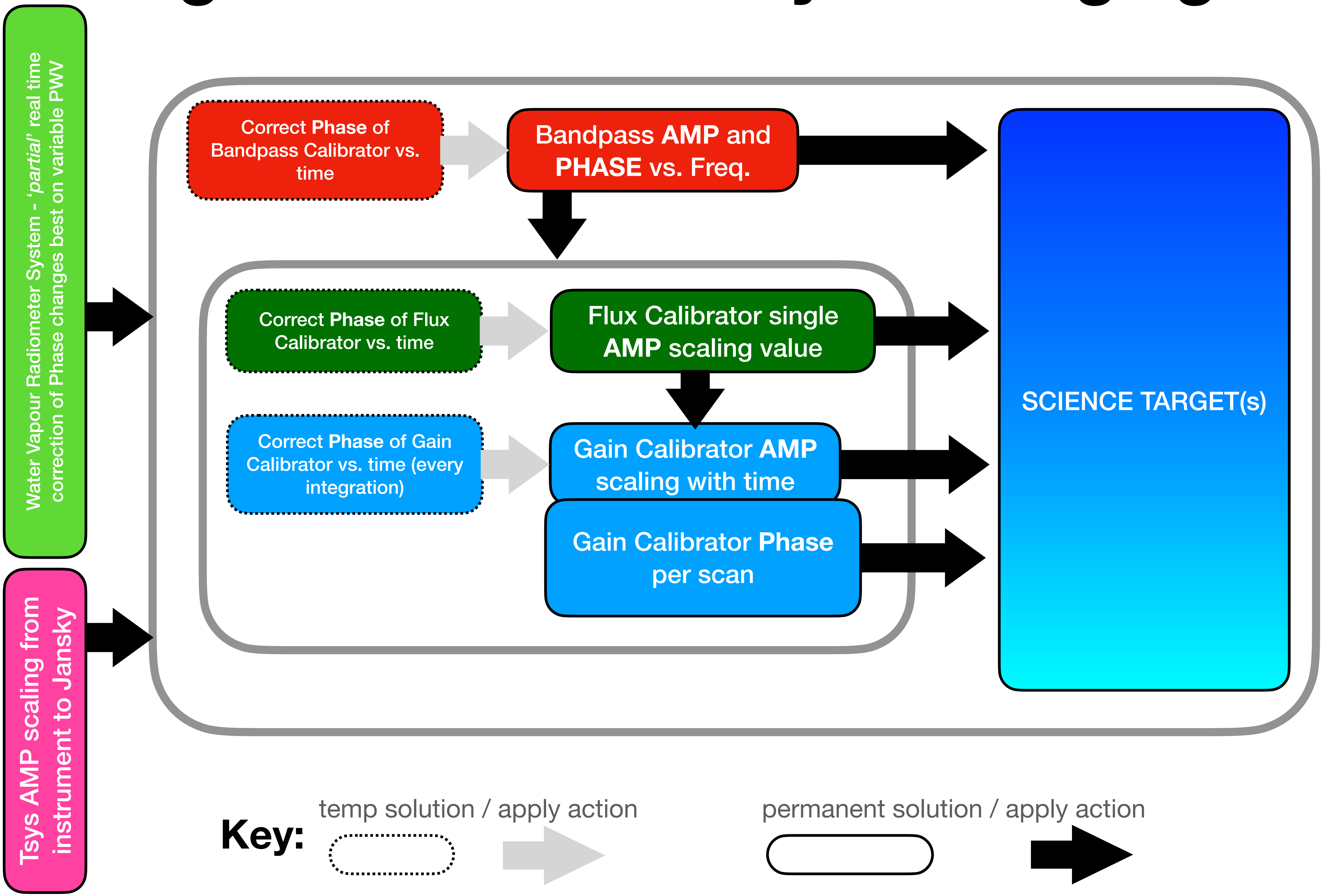
Part 1

Data get calibrated ready for imaging



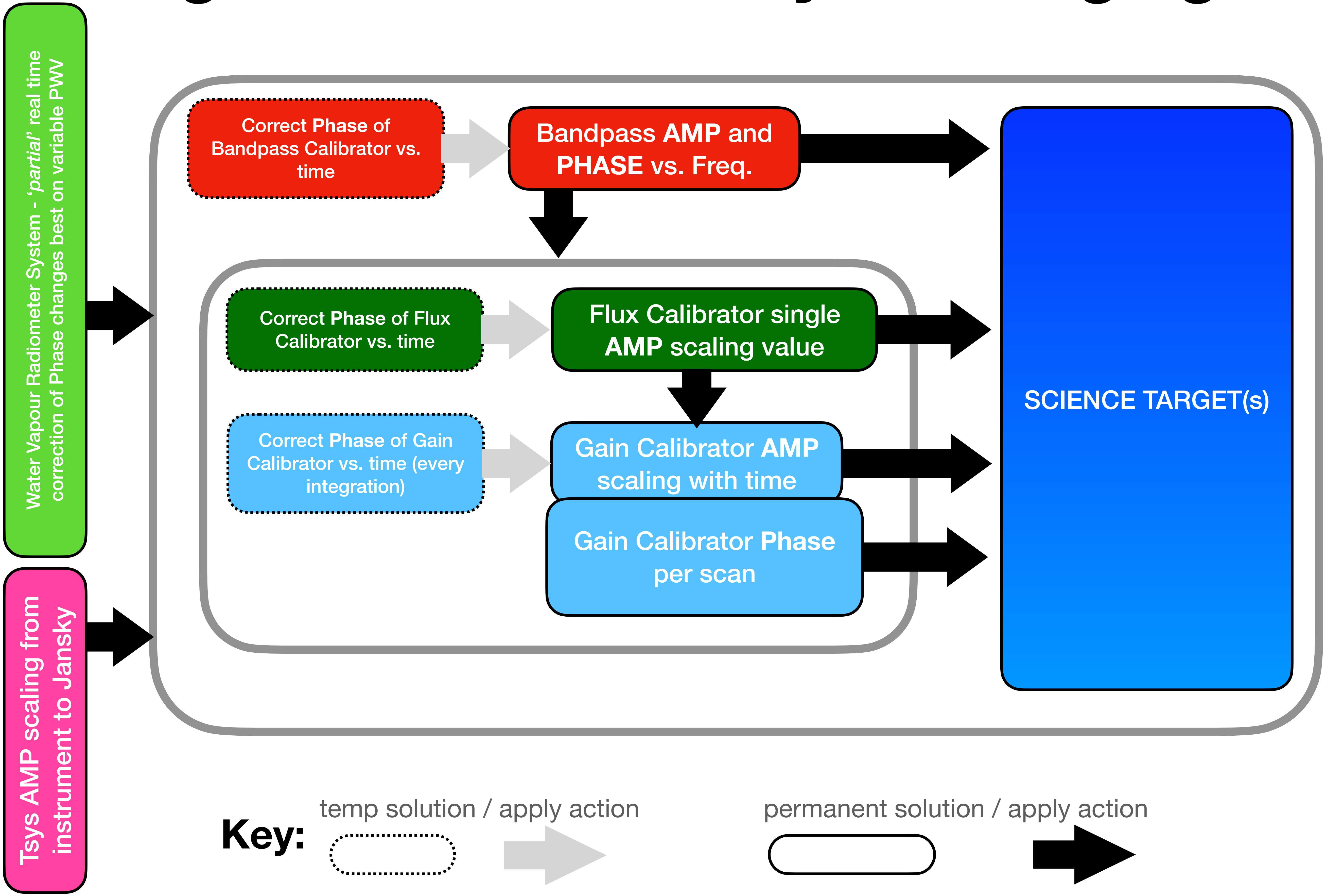
Part 1

Data get calibrated ready for imaging



Part 1

Data get calibrated ready for imaging



Any questions at this stage?

- **PART 2 : What is in your ALMA data**
  - what do you get
  - how is everything organised
  - how can you look at things

Part 2

Download from Archive

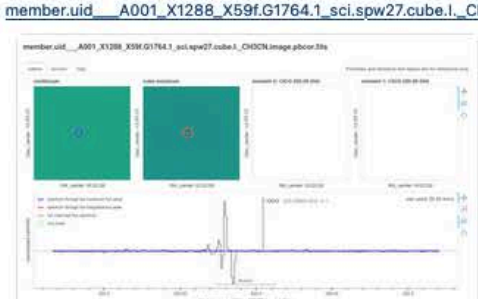
you'll also want the ASDM (big) to restore and image the data

Download 10 GB | Open legacy Request Handler

- Project (1)
- Group ObsUniSet (1)
- Member ObsUniSet (1)
- Source (1)
- Collection (1)
- Array (1)
- File type (9)
- File class (12)

☑ Select all ☑ Readme ☑ Product tar ☑ Auxiliary tar ☐ Raw tgz ☐ Raw (semipass) tgz ☐ Extern

	Name		Size	↑ Project	↑ GOUS	↑ MOUS
☐	member.uid___A001_X1288_X59f.casa-20171206-075656.log	(auxiliary, log)	419 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	2017.1.00098.S_uid___A002_Xc55c89_X120.asdm.sdm.tar	(raw)	127 GB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.casa-20171221-072759.log	(auxiliary, log)	625 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.scriptForImaging.py	(auxiliary, script)	14 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.casa_pipescript.py	(auxiliary, script)	2 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.J1830-1606_chk.spw31.mfs.I.pb.fits.gz	(product)	21 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☑	uid___A002_Xc55c89_X120.ms.flagversions.tgz	(auxiliary, calibration)	3 MB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	uid___A002_Xc55c89_X120.qa0_report.pdf	(auxiliary, qa)	670 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☑	member.uid___A001_X1288_X59f.README.txt	(readme)	3 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.qa2_report.pdf	(auxiliary, qa)	64 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.antennapos.csv	(auxiliary, calibration)	2 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☑	member.uid___A001_X1288_X59f.session_2.caltables.tgz	(auxiliary, calibration)	18 MB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.flux.csv	(auxiliary, calibration)	2 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.J1830-1606_chk.spw25.mfs.I.pb.fits.gz	(product)	20 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.qa2_report.html	(auxiliary, qa)	68 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☑	member.uid___A001_X1288_X59f.hifa_calimage.weblog.tgz	(auxiliary, qa)	223 MB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.casa-20171107-105701.log	(auxiliary, log)	553 B	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.casa-20171221-163345.log	(auxiliary, log)	661 B	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☑	2017.1.00098.S_uid___A001_X1288_X59f_auxiliary.tar	(auxiliary)	262 MB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.G1764.1_sci.spw27.cube.I._CH3CN.flux.fits.gz	(product)	250 MB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.casa-20171107-110233.log	(auxiliary, log)	553 B	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.casa-20171124-143154.log	(auxiliary, log)	354 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.G1764.1_sci.spw27.cube.I._CH3CN.image.pbcor.fits	(product)	4 GB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f

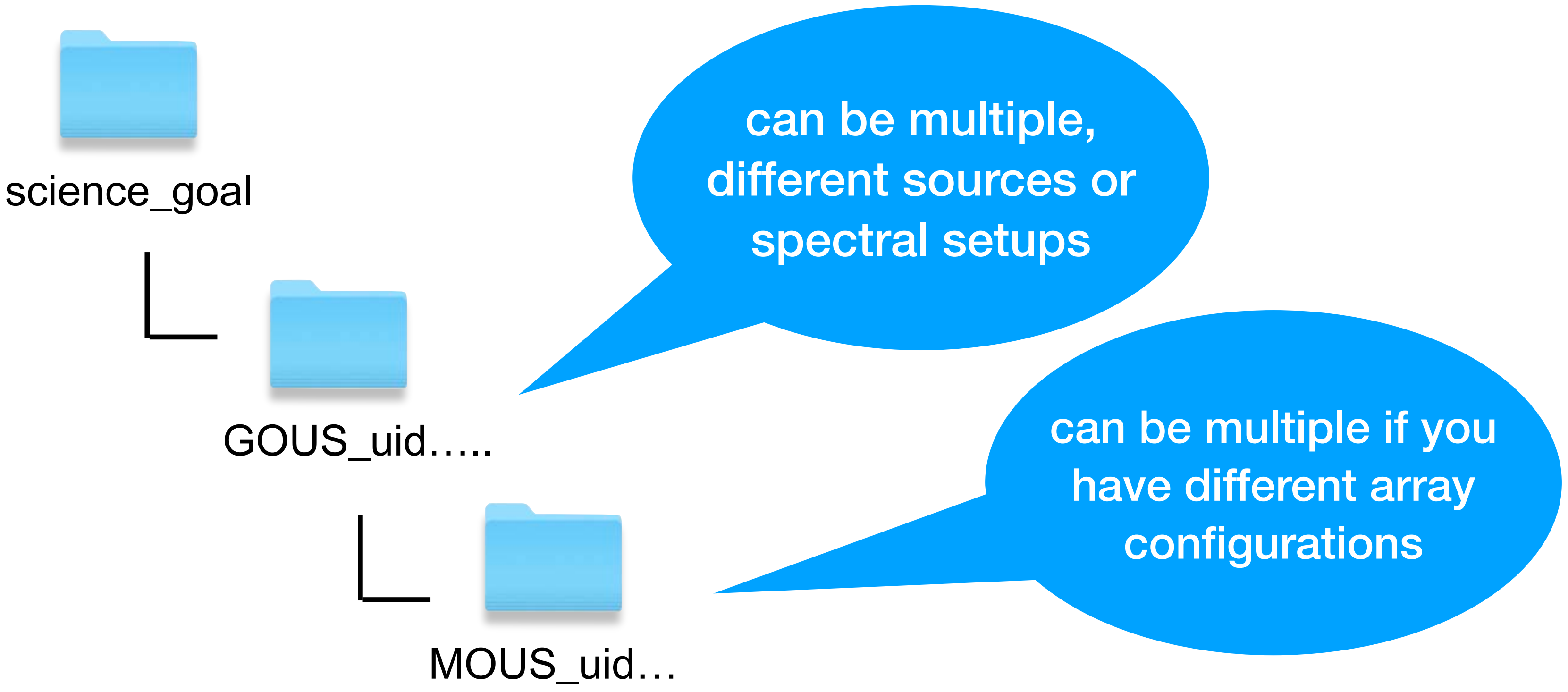


Band: 6
Frequency range: 218.854..220.728
Frequency resolution: 1,128.906 kHz
Line sens. (10km/s): 0.473mJy/beam
Line sens. (native): 0.03uJy/beam
Polaritazions: XX YY
Array: 12m

☐	member.uid___A001_X1288_X59f.casa-20171213-153436.log	(auxiliary, log)	4 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☑	member.uid___A001_X1288_X59f.G1764.1_sci.spw27.cube.I._CH3CN.mask.tgz	(product)	4 MB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.J1825-1718_ph.spw25.mfs.I.pb.fits.gz	(product)	20 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.hifa_calimage.casa_commands.log	(auxiliary, log)	88 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.calimage.pipeline_manifest.xml	(auxiliary, script)	4 kB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.calimage.product_rename.txt	(auxiliary, script)	3 B	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f
☐	member.uid___A001_X1288_X59f.G1764.1_sci.spw25.cube.I._SiO.image.pbcor.fits	(product)	5 GB	2017.1.00098.S	uid://A001/X1288/X59e	uid://A001/X1288/X59f

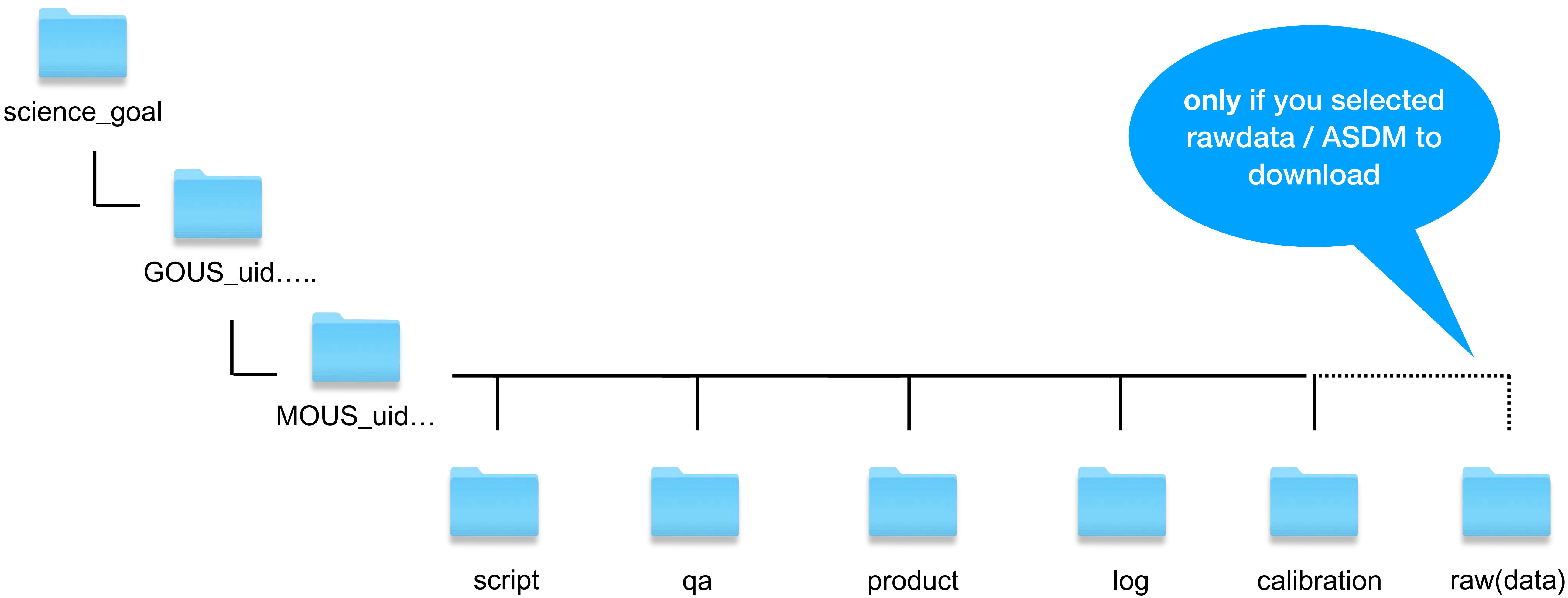
Part 2

- **Folder breakdown : From the Science Goal**
  - Group Observation Unit Set (GOUS)
  - Member Observation Unit Set (MOUS)

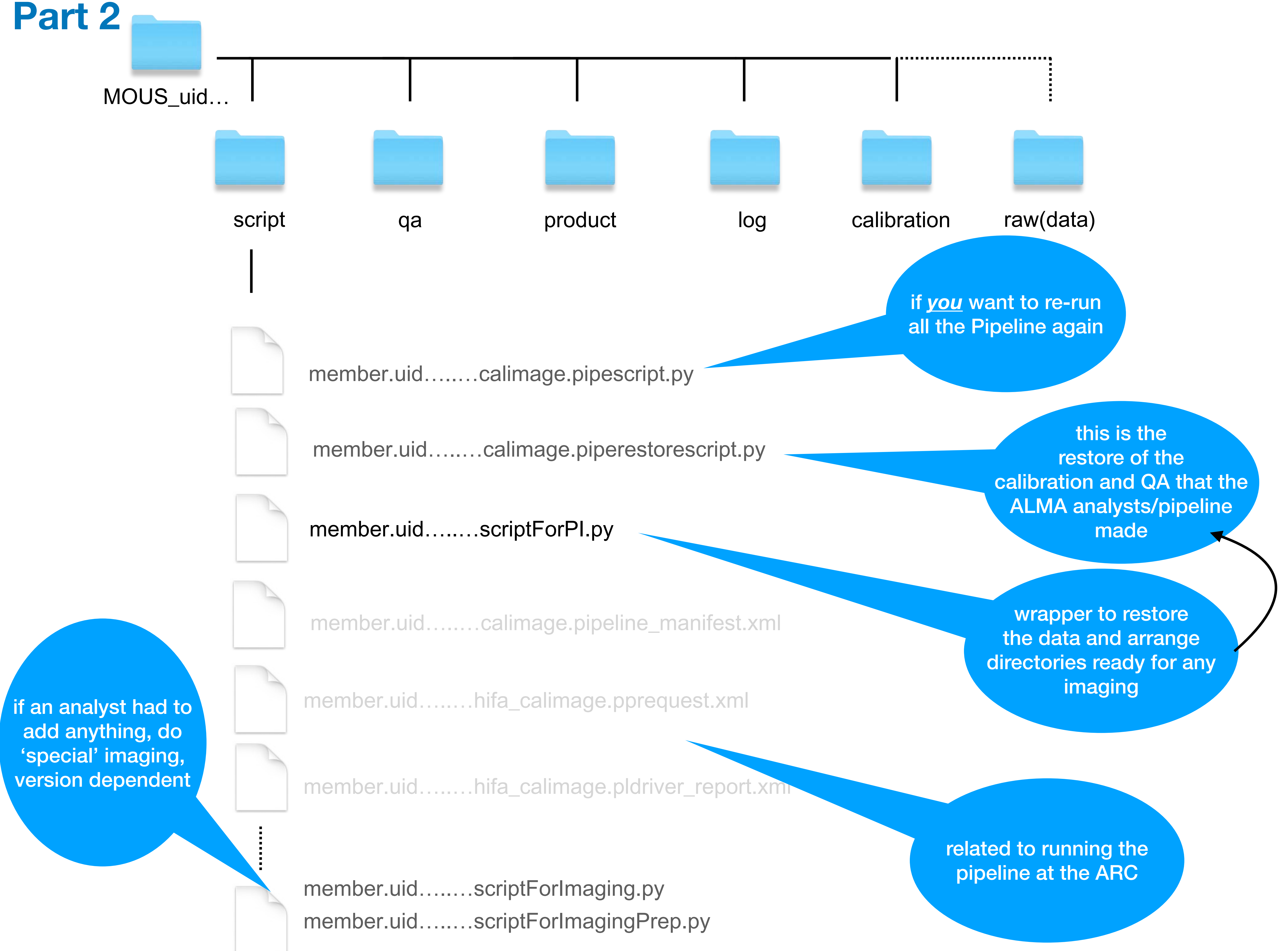


Part 2

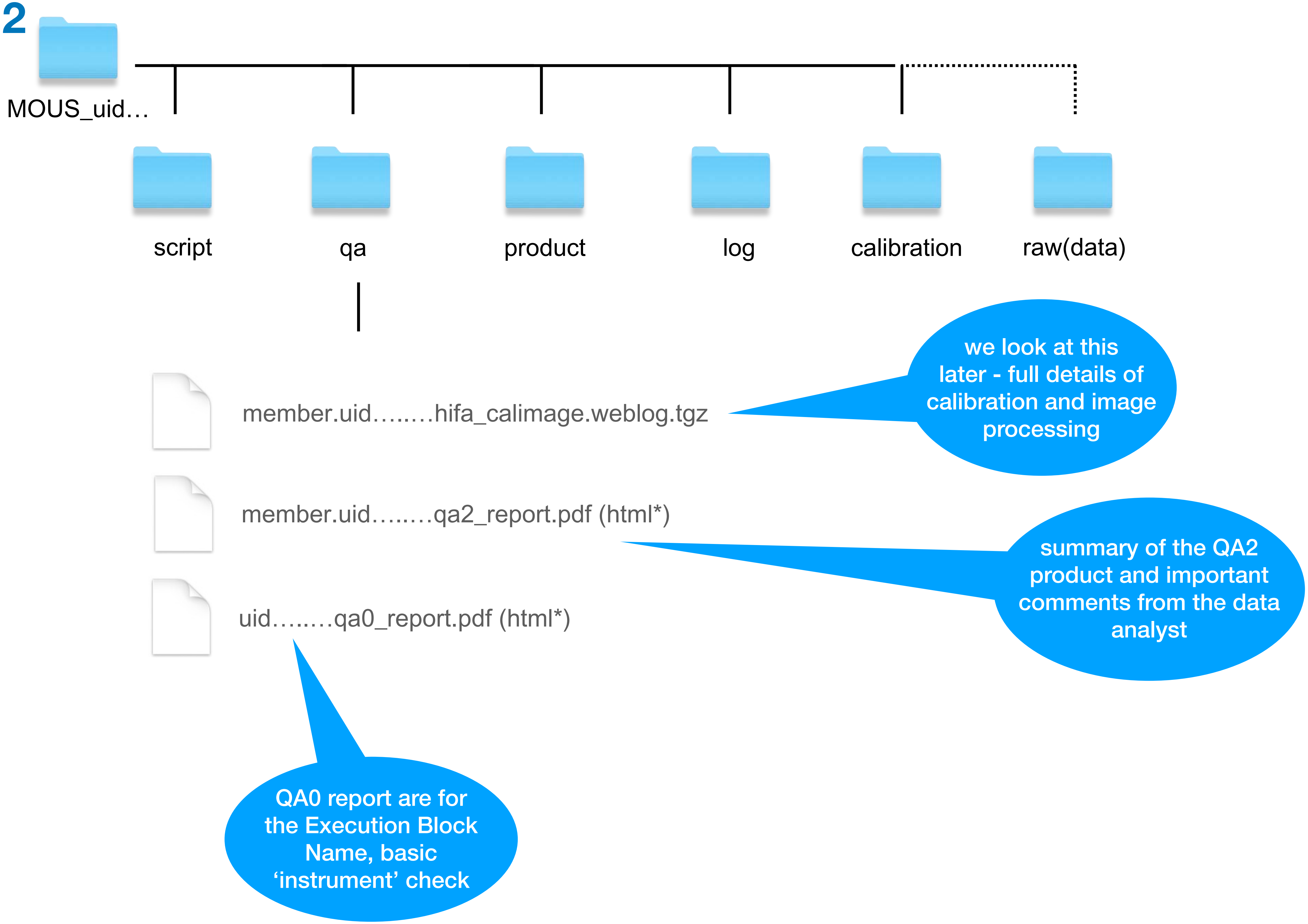
- **Folder breakdown : From the Science Goal**
  - Member Observation Unit Set (MOUS) ➡ 'working area'
  - all directories inc, **raw data** stores the ASDM (ALMA science data model),



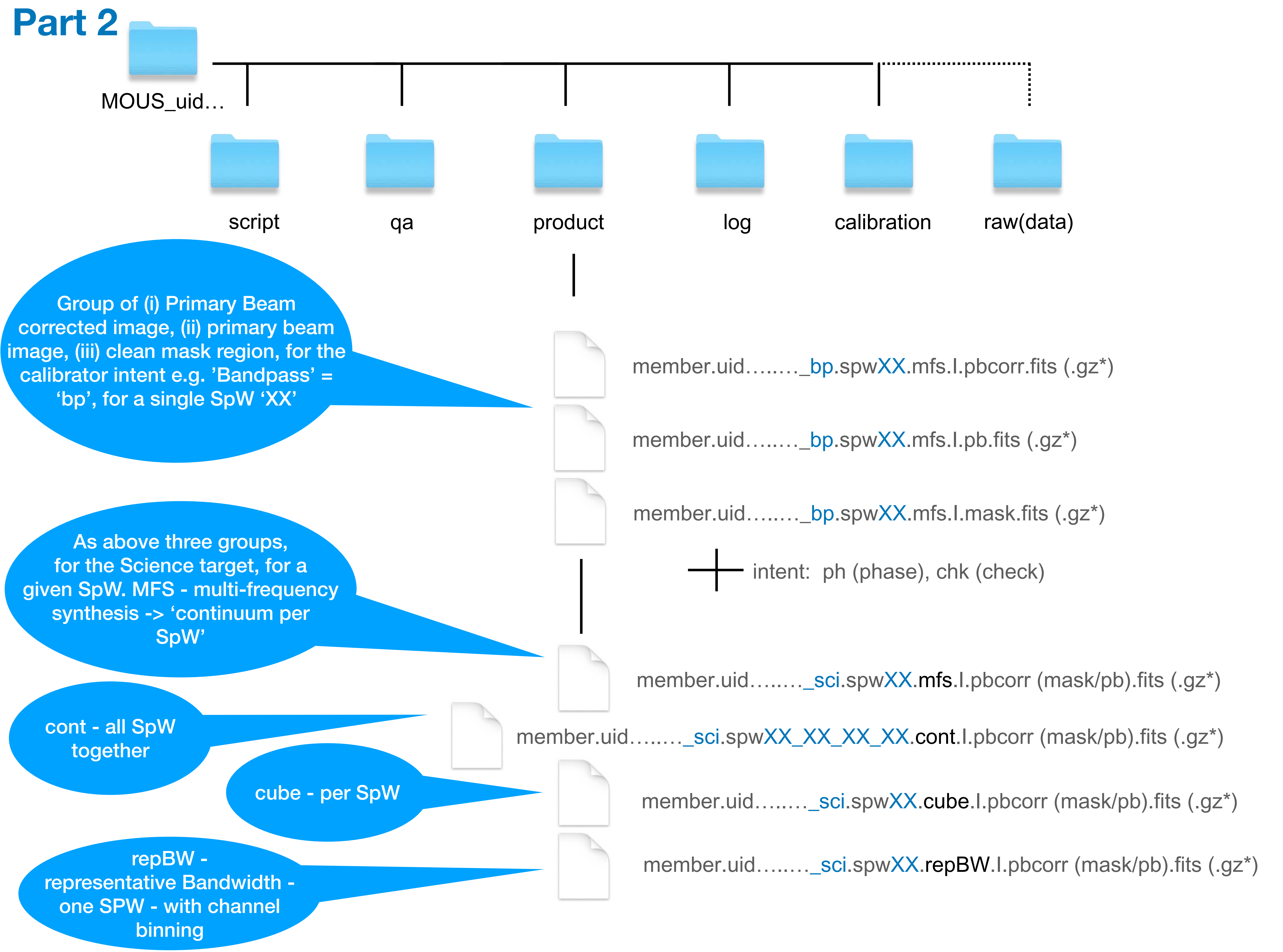
Part 2


MOUS_uid…
script
qa
product
log
calibration
raw(data)
member.uid……..calimage.pipescript.py
member.uid……..calimage.piperestorescript.py
member.uid……..scriptForPI.py
member.uid……..calimage.pipeline_manifest.xml
member.uid……..hifa_calimage.pprequest.xml
member.uid……..hifa_calimage.pldriver_report.xml
member.uid……..scriptForImaging.py
member.uid……..scriptForImagingPrep.py
if *you* want to re-run all the Pipeline again
this is the restore of the calibration and QA that the ALMA analysts/pipeline made
wrapper to restore the data and arrange directories ready for any imaging
if an analyst had to add anything, do ‘special’ imaging, version dependent
related to running the pipeline at the ARC

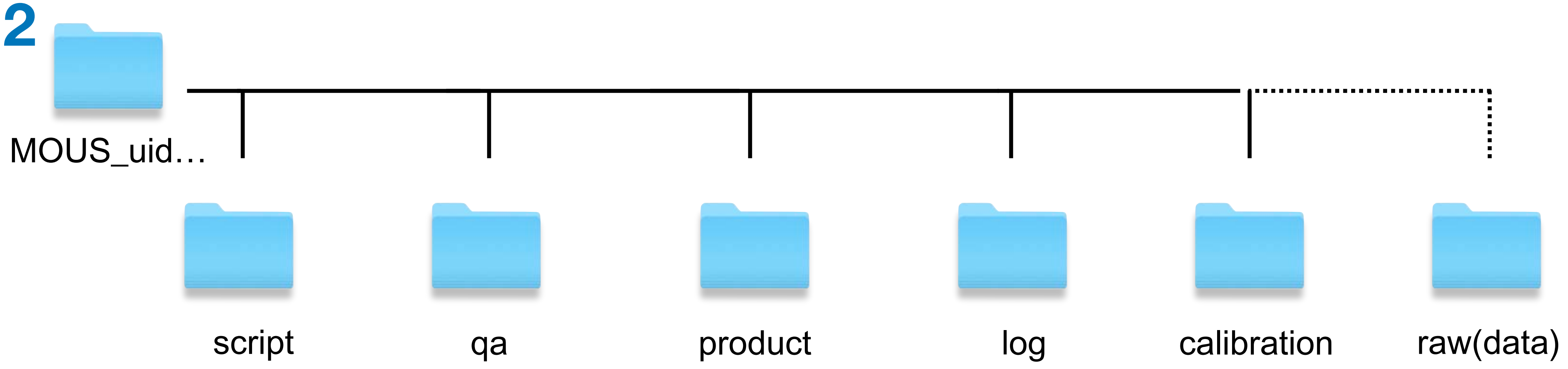
Part 2

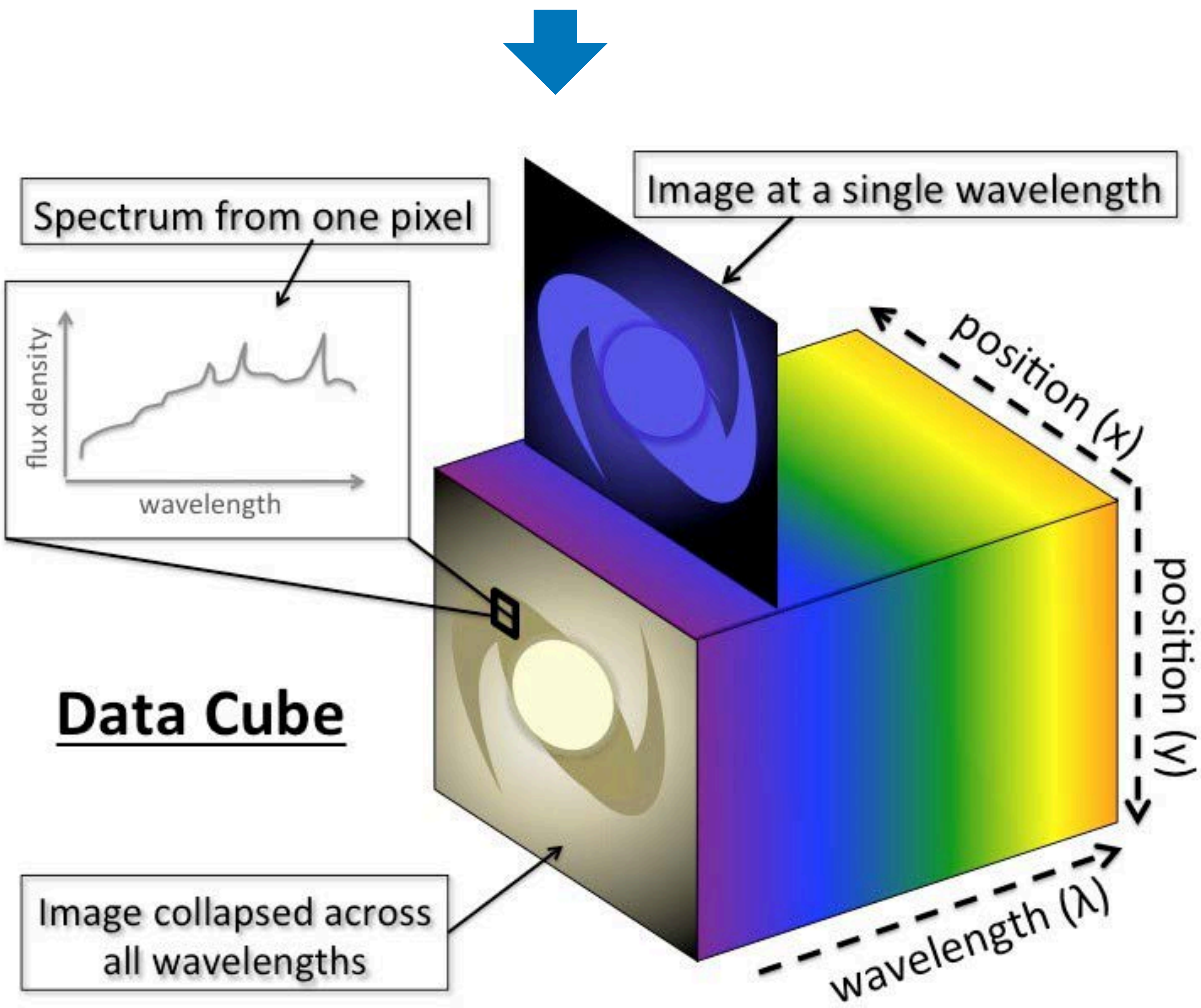


*version dependent (CASA/Pipeline)

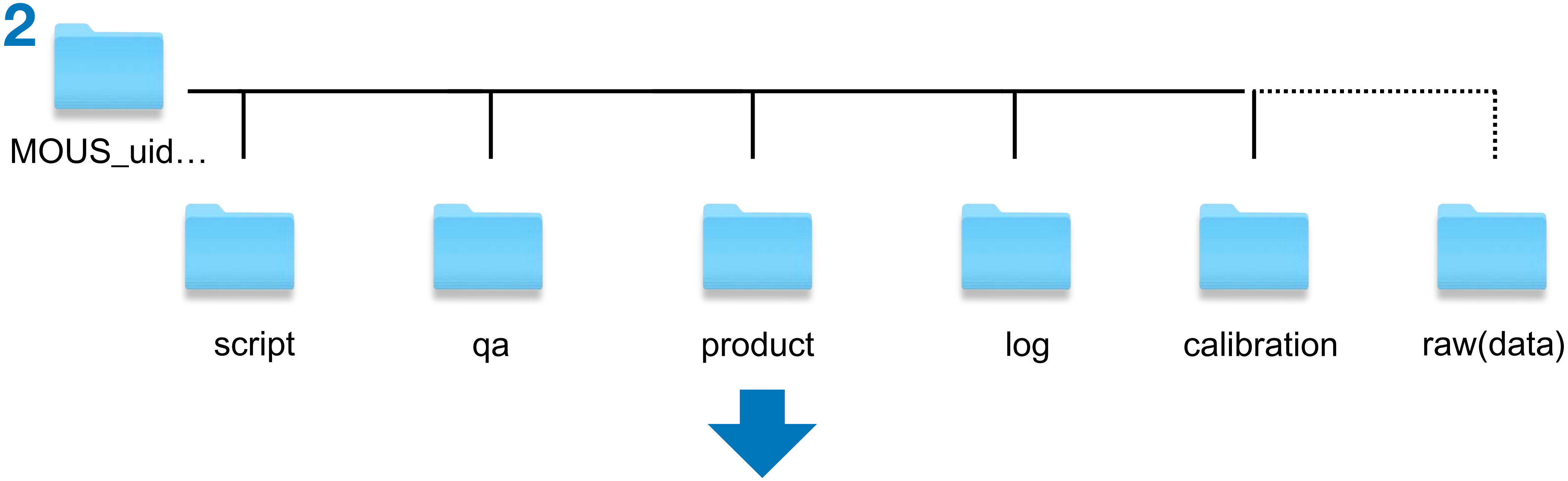

Part 2
MOUS_uid…
script
qa
product
log
calibration
raw(data)
Group of (i) Primary Beam corrected image, (ii) primary beam image, (iii) clean mask region, for the calibrator intent e.g. 'Bandpass' = 'bp', for a single SpW 'XX'
member.uid…….._bp.spwXX.mfs.I.pbcorr.fits (.gz*)
member.uid…….._bp.spwXX.mfs.I.pb.fits (.gz*)
member.uid…….._bp.spwXX.mfs.I.mask.fits (.gz*)
intent: ph (phase), chk (check)
As above three groups, for the Science target, for a given SpW. MFS - multi-frequency synthesis -> 'continuum per SpW'
member.uid…….._sci.spwXX.mfs.I.pbcorr (mask/pb).fits (.gz*)
cont - all SpW together
member.uid…….._sci.spwXX_XX_XX_XX.cont.I.pbcorr (mask/pb).fits (.gz*)
cube - per SpW
member.uid…….._sci.spwXX.cube.I.pbcorr (mask/pb).fits (.gz*)
repBW - representative Bandwidth - one SPW - with channel binning
member.uid…….._sci.spwXX.repBW.I.pbcorr (mask/pb).fits (.gz*)

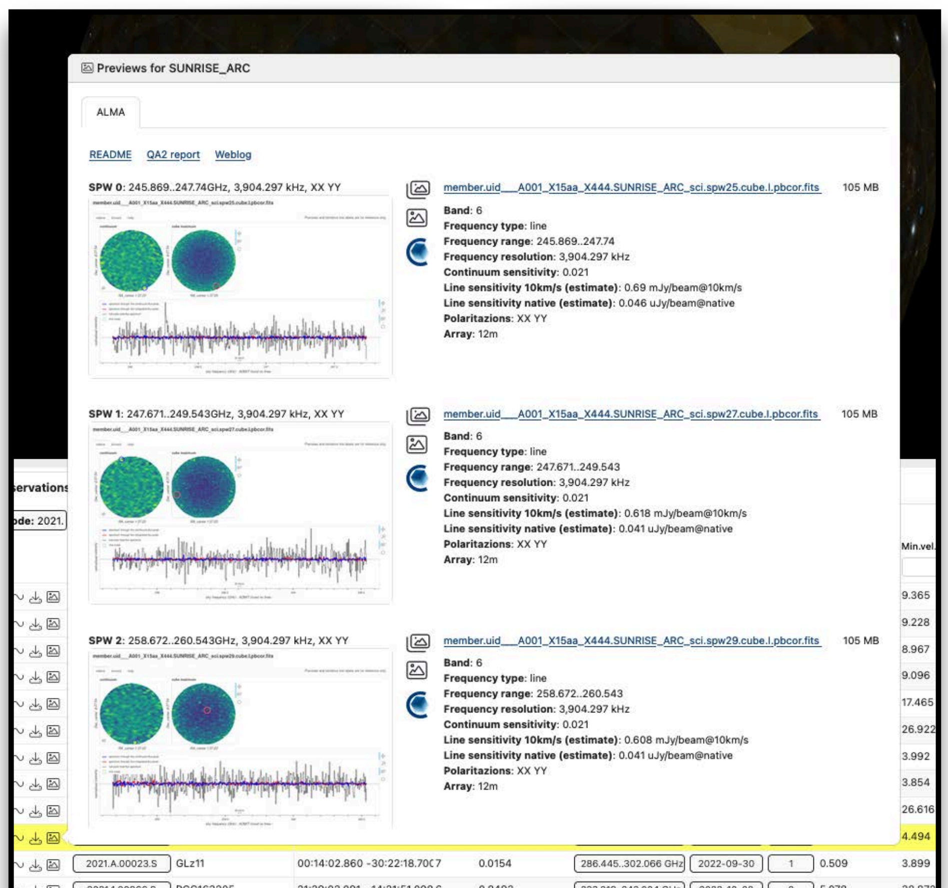
Part 2


MOUS_uid…
script
qa
product
log
calibration
raw(data)


Spectrum from one pixel
flux density
wavelength
Image at a single wavelength
position (x)
position (y)
wavelength (λ)
Data Cube
Image collapsed across all wavelengths

Part 2

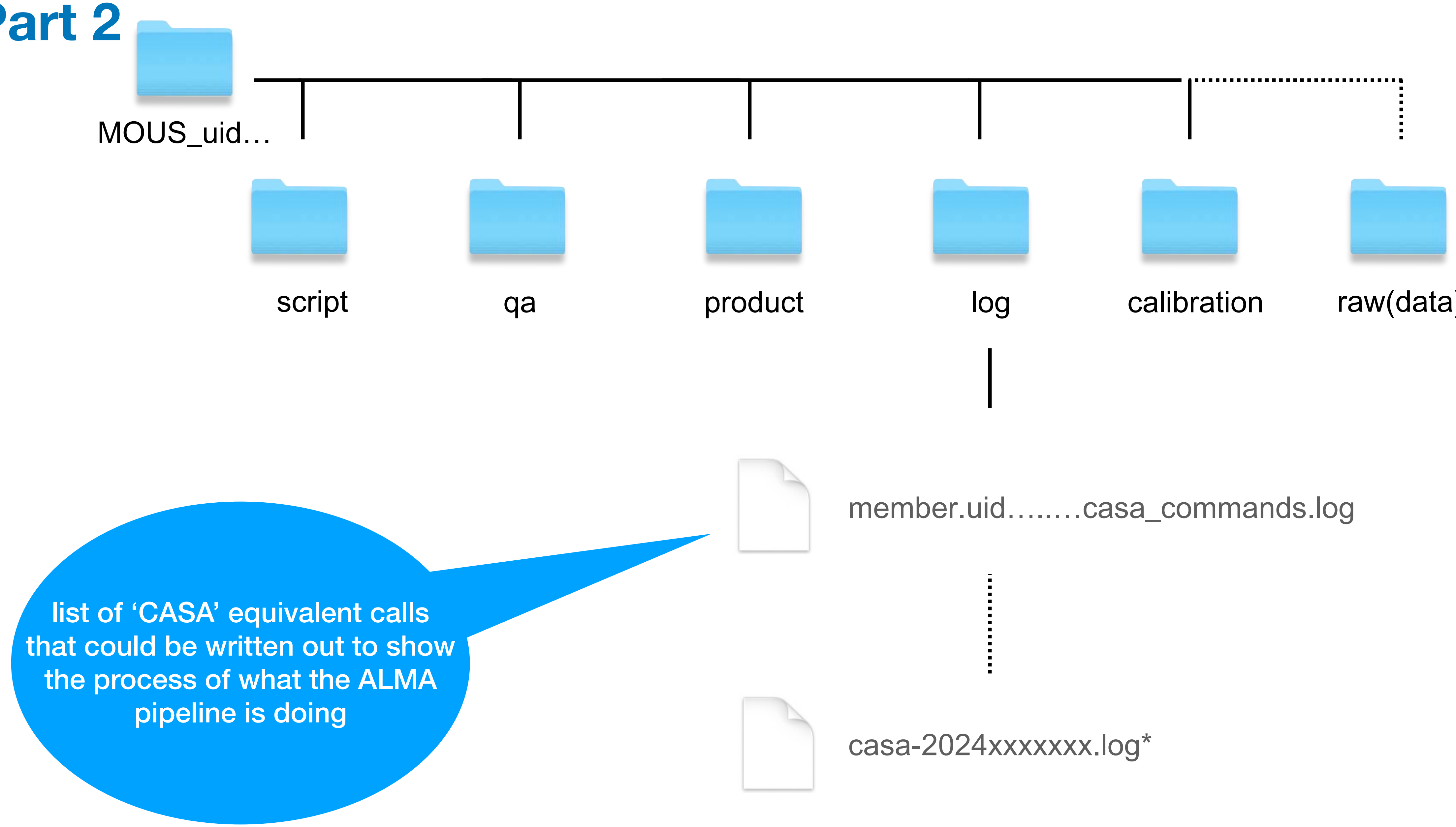




ALMA Archive (preview)

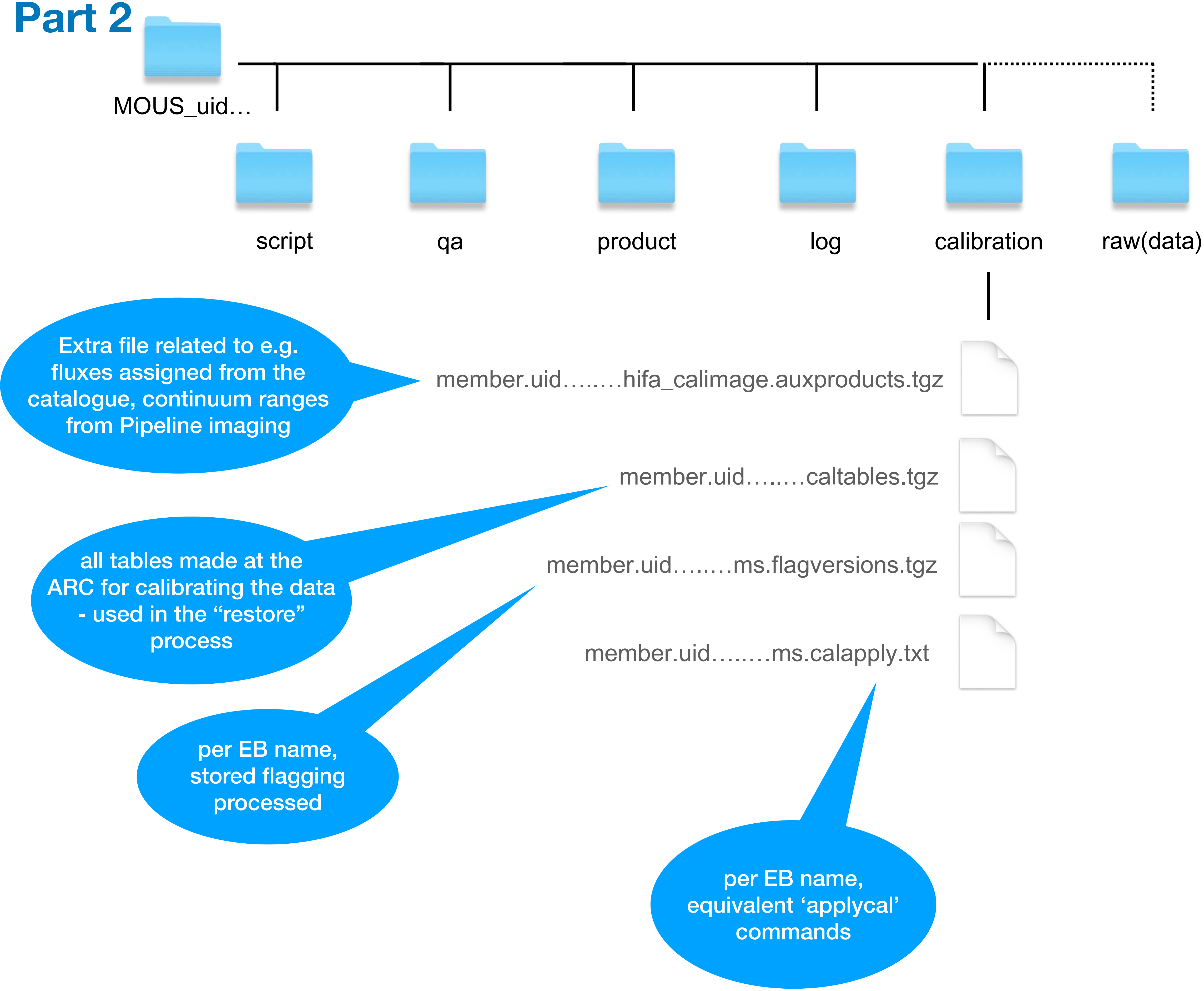
CARTA

Part 2

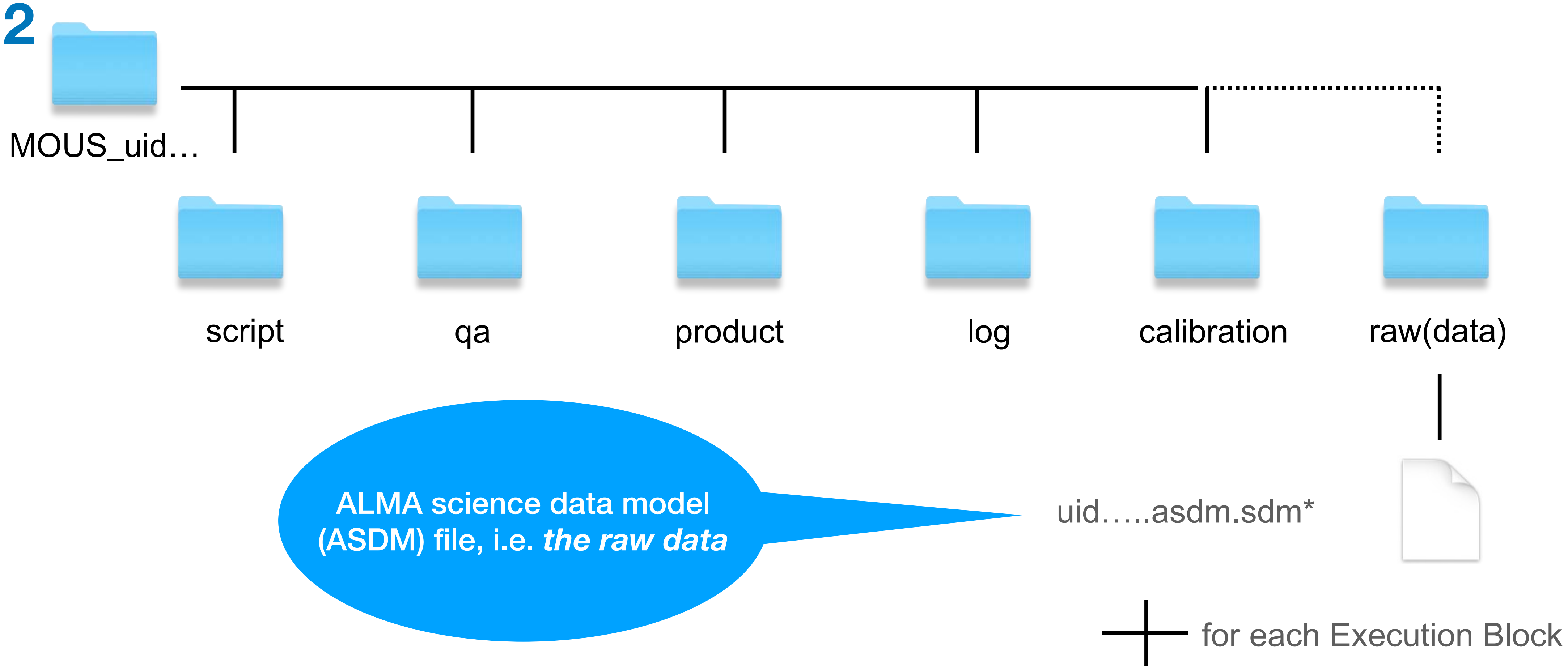


*version dependent (CASA/Pipeline)

Part 2

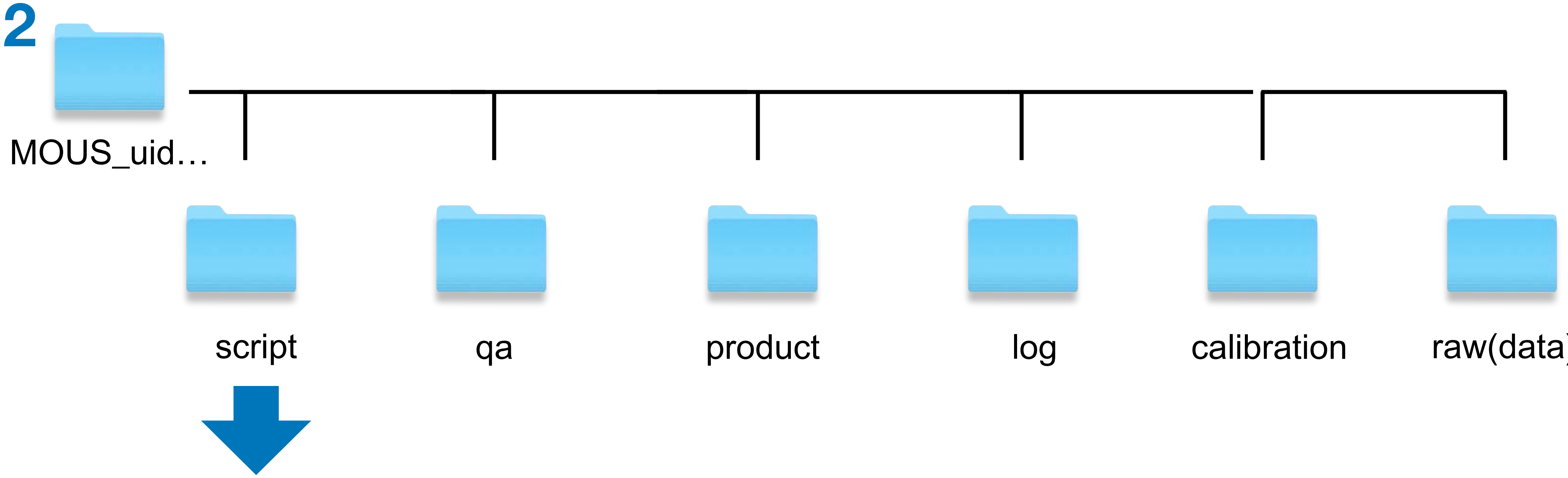

MOUS_uid…
script
qa
product
log
calibration
raw(data)
member.uid…….hifa_calimage.auxproducts.tgz
member.uid…….caltables.tgz
member.uid…….ms.flagversions.tgz
member.uid…….ms.calapply.txt
Extra file related to e.g. fluxes assigned from the catalogue, continuum ranges from Pipeline imaging
all tables made at the ARC for calibrating the data - used in the "restore" process
per EB name, stored flagging processed
per EB name, equivalent 'applycal' commands

Part 2



*could be a tar at download, but downloader wrapper untars this

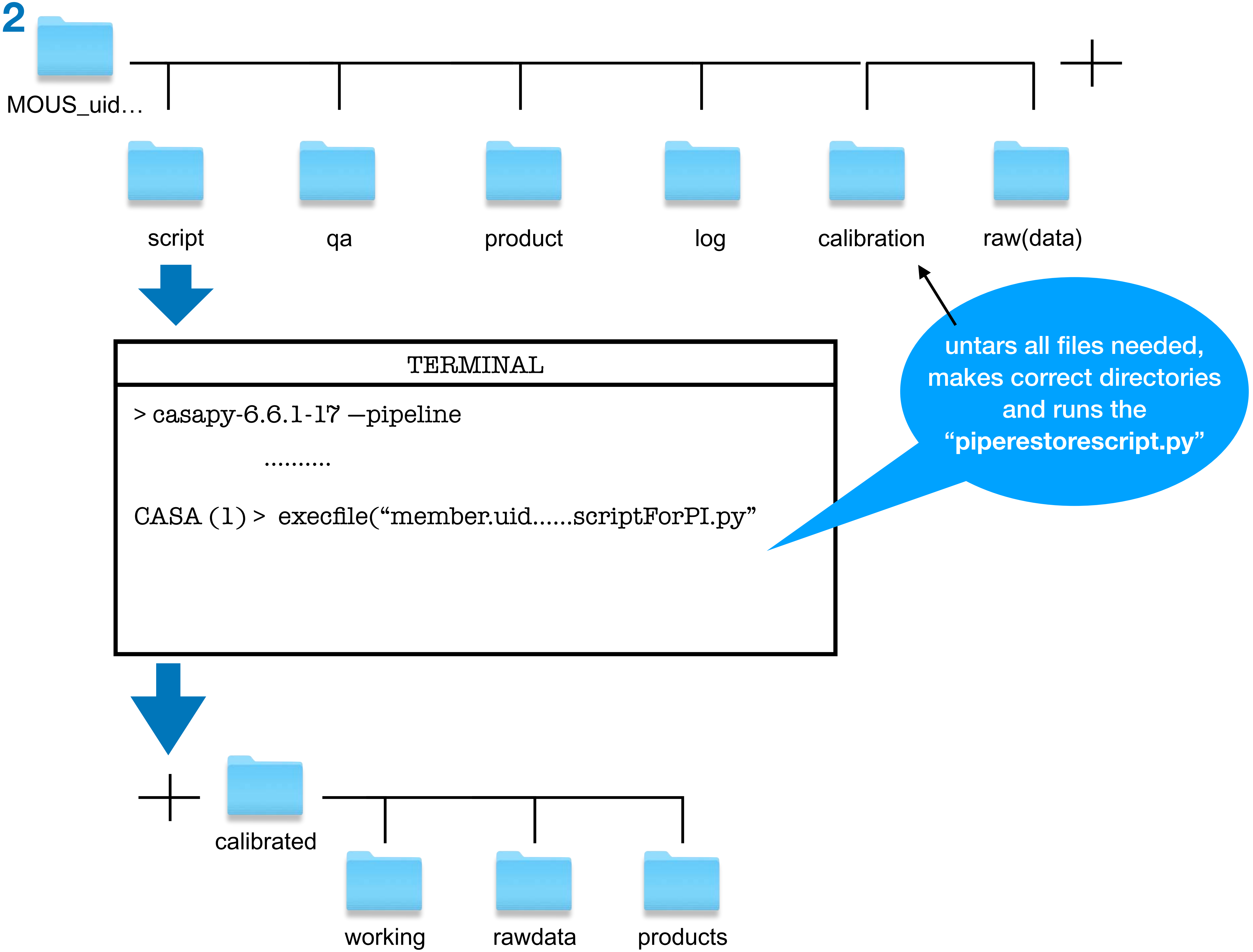
Part 2



TERMINAL

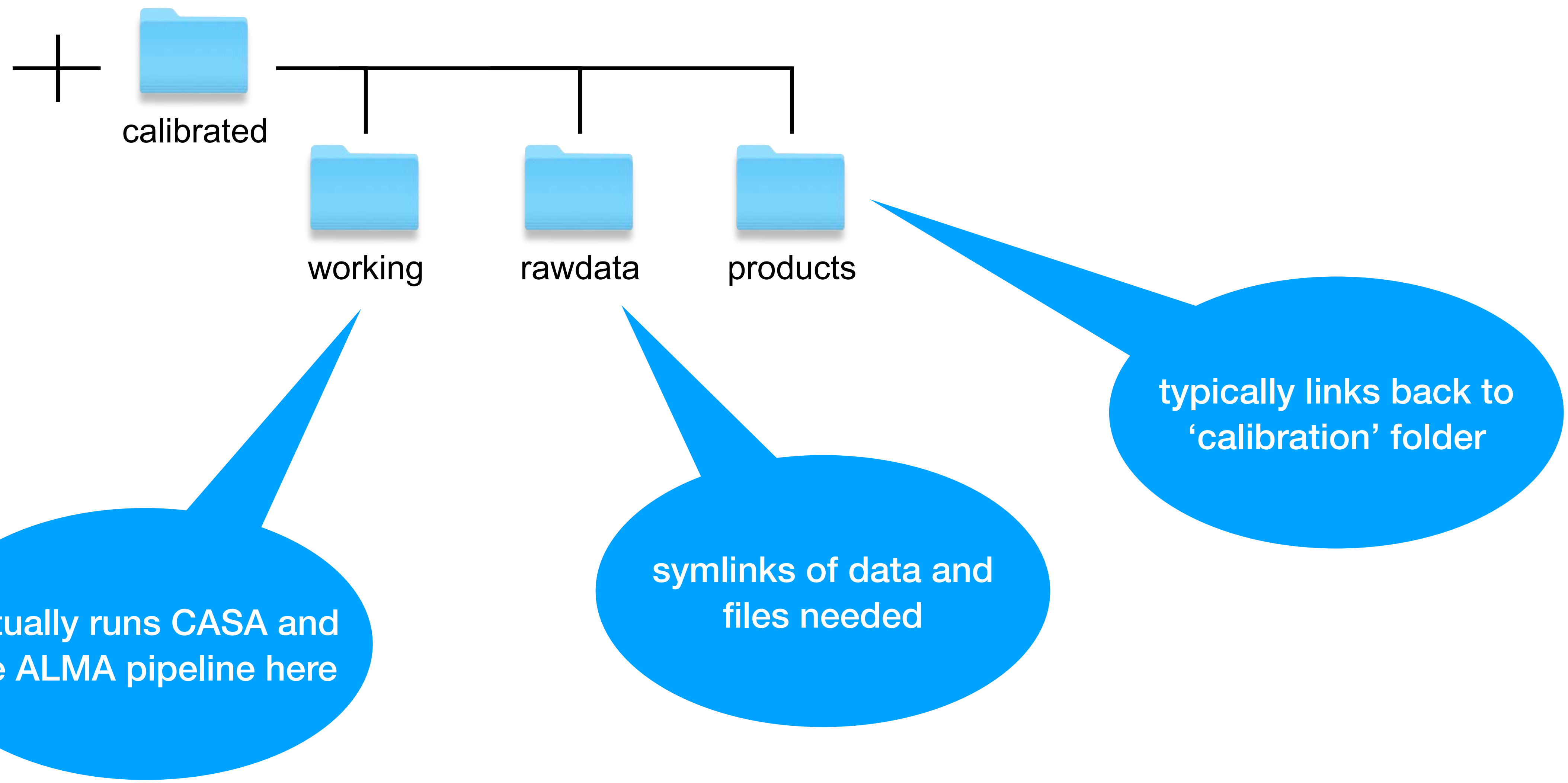
```
> casapy-6.6.1-17 —pipeline

2025-02-13 13:11:04 INFO: Environment is not MPI enabled. Pipeline operating in single host mode
2025-02-13 13:11:05 INFO: Environment variable FLUX_SERVICE_URL not defined.  Switching to backup url.
2025-02-13 13:11:05 INFO: Environment variable FLUX_SERVICE_URL_BACKUP not defined.
2025-02-13 13:11:05 INFO: Pipeline version 2024.1.0.8 running on arcp17.hq.eso.org
2025-02-13 13:11:05 INFO: Host environment:
        CPU: Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz (physical cores: 8, logical cores: 16)
        Memory: 503.5 GiB RAM, 8.0 GiB swap
        OS: Red Hat Enterprise Linux 8.6 (Ootpa)
        cgroup limits: 100% of 8 CPU cores, memory limits=244.1 GiB
        ulimit limits: CPU time=N/A, memory=262144000000, files=131072
2025-02-13 13:11:05 INFO: Environment as detected by CASA:
        CPUs reported by CASA: 8 cores, max 8 OpenMP threads
        Available memory: 244.1 GiB
2025-02-13 13:11:05 INFO: Initializing cli...
2025-02-13 13:11:05 INFO: Loaded Pipeline commands from package: h
2025-02-13 13:11:05 INFO: Loaded Pipeline commands from package: hif
2025-02-13 13:11:05 INFO: Loaded Pipeline commands from package: hifa
2025-02-13 13:11:05 INFO: Loaded Pipeline commands from package: hifv
2025-02-13 13:11:05 INFO: Loaded Pipeline commands from package: hsd
2025-02-13 13:11:05 INFO: Loaded Pipeline commands from package: hsdn
*** startup.py: QA2-relevant modules will be imported ***
casaVersion =  6.6.1.17
scipy version =  1.10.1
analysisUtils.py: imported casatasks and casatools individually
Using astropy.io.fits instead of pyfits
$Id: analysisUtils.py,v 2.248 2025/02/04 18:31:12 thunter Exp $
CASA 6.6.1.17 -- Common Astronomy Software Applications [6.6.1.17]
```

Part 2



MOUS_uid…
script
qa
product
log
calibration
raw(data)
TERMINAL
> casapy-6.6.1-17 —pipeline
……….
CASA (1) > execfile("member.uid……scriptForPI.py"
untars all files needed, makes correct directories and runs the "piperestorescript.py"
calibrated
working
rawdata
products

Part 2



You now have “uid…..ms” (or also .cal) these are calibrated data and can be imaged - with Pipeline or CASA ‘tclean’

Part 2

Requesting calibrated data in Europe

from 1 October 2019 onwards

By popular demand, the EU ARC has implemented a service which permits ALMA users to request the calibrated data for a given dataset (Member Obs Unit Set, MOUS) to be made available for download. The service is open both for ALMA PIs or Delegees with proprietary ALMA data and for archival users wanting to use datasets for which the proprietary time has expired.

If you have identified a particular MOUS that you want to investigate, please file a normal Helpdesk ticket in the department "Archive and Data Retrieval (EU)" and select the "Data request" sub-category.

In the body of the text always specify the project code (e.g. 2015.1.09999.S, one per ticket) and MOUS UID(s) (e.g. uid___A001_X340_X6 or uid://A001/X340/X6). You can enumerate up to 10 MOUSs in your request.

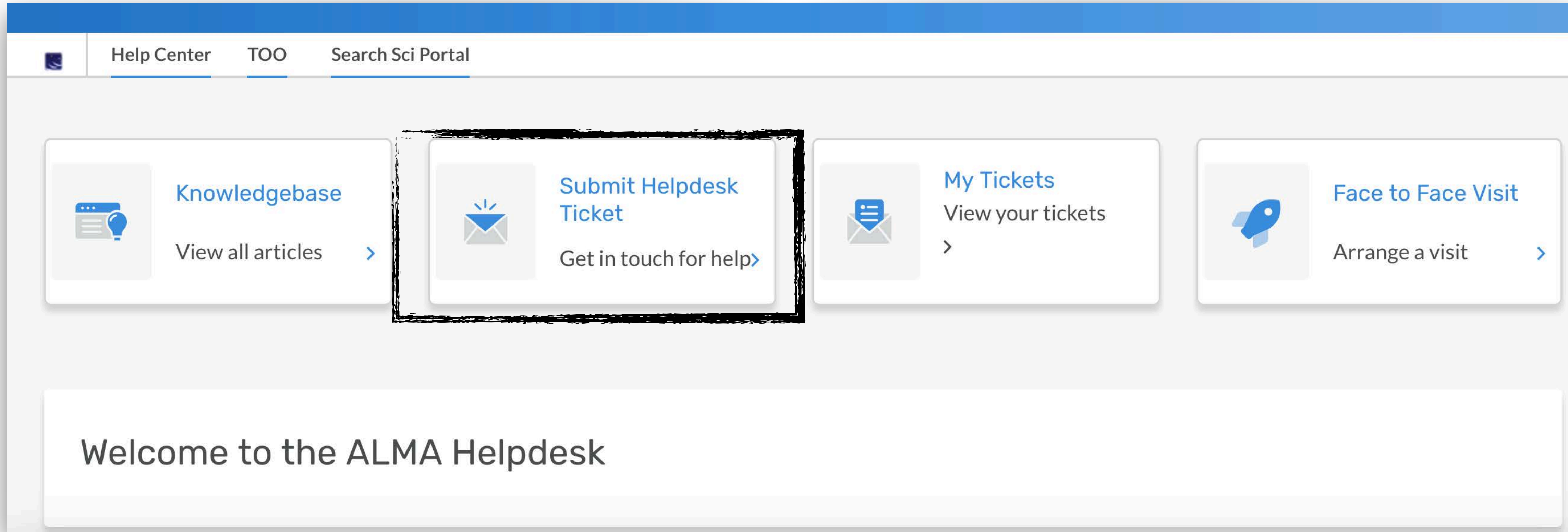
The creation and staging will be done one MOUS at a time and you will be notified *by separate email for each MOUS* as to where you can download the tarred MS(s). Depending on the workload on the EU ARC systems, it may take days before your dataset is ready for you.

Your download link will remain valid for 28 days, which means that you have 28 days from the time of the notification email to download the data.

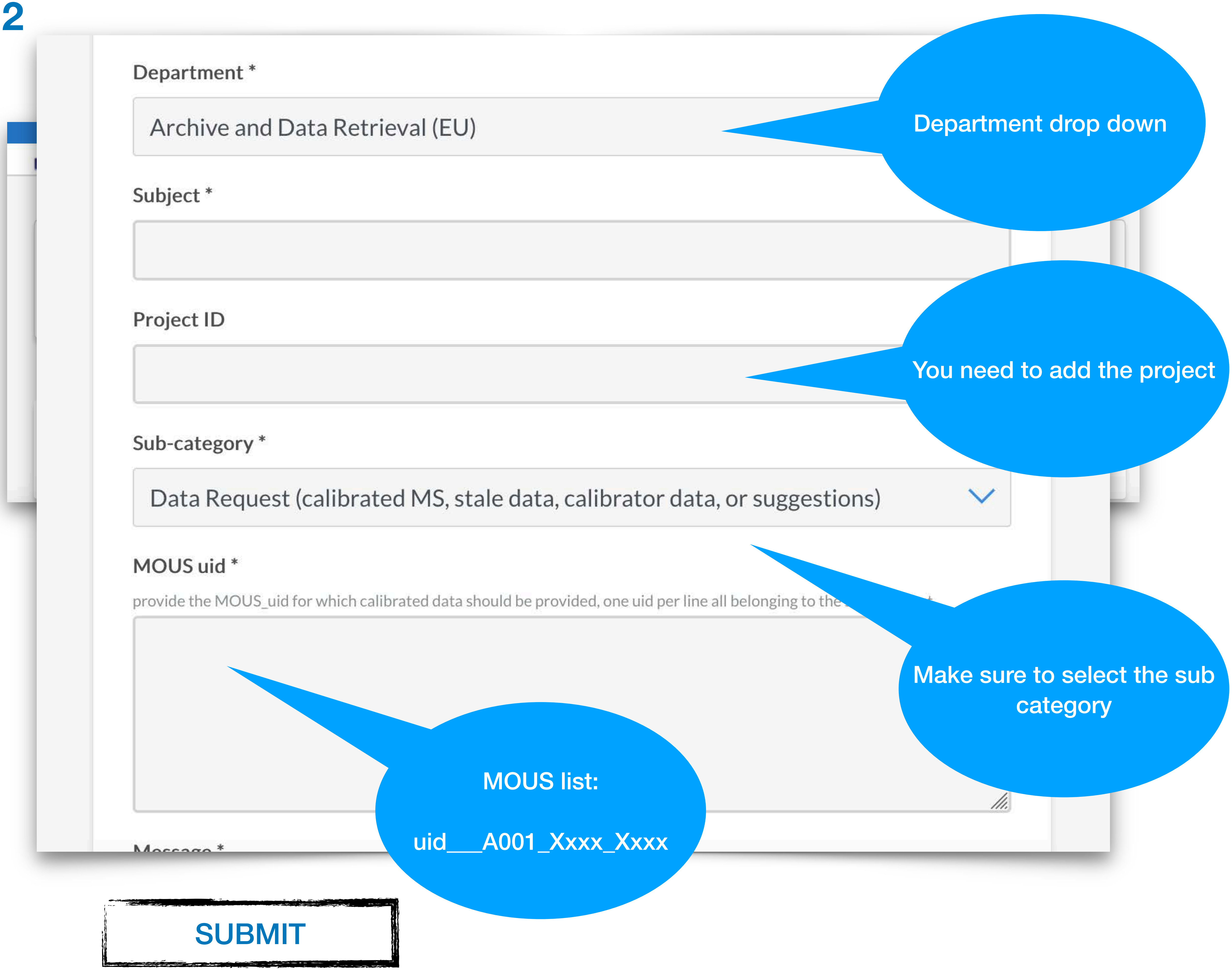
The service will become available on 1 October 2019.

can also request from the EU ARC to calibrated the data for you

Part 2


Help Center
TOO
Search Sci Portal
Knowledgebase
View all articles
Submit Helpdesk Ticket
Get in touch for help
My Tickets
View your tickets
Face to Face Visit
Arrange a visit
Welcome to the ALMA Helpdesk

Part 2


Department *
Archive and Data Retrieval (EU)
Subject *
Project ID
Sub-category *
Data Request (calibrated MS, stale data, calibrator data, or suggestions)
MOUS uid *
provide the MOUS_uid for which calibrated data should be provided, one uid per line all belonging to the
Department drop down
You need to add the project
Make sure to select the sub category
MOUS list:
uid___A001_Xxxx_Xxxx
SUBMIT

MORE READING:

- **ALMA documents -** almascience.org/documents-and-tools
- **ALMA technical handbook / proposers guide (above)**
- **Archive Primer -** almascience.org/documents-and-tools/cycle11/archive-primer
- **Online "I-Train" tutorials -** almascience.eso.org/tools/eu-arc-network/i-train

  - **Interferometry Schools:**
    - NRAO Synthesis imaging workshops
    - ERIS European Radio Interferometry Schools
    - IRAM Interferometry schools

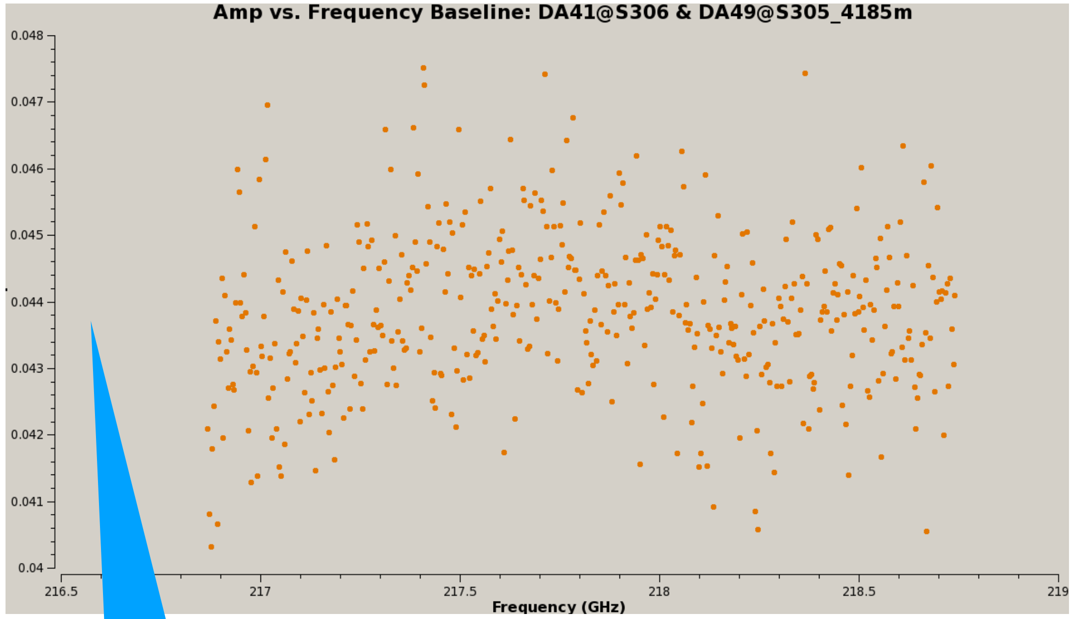
THANK YOU - any questions ???

Extra Slides

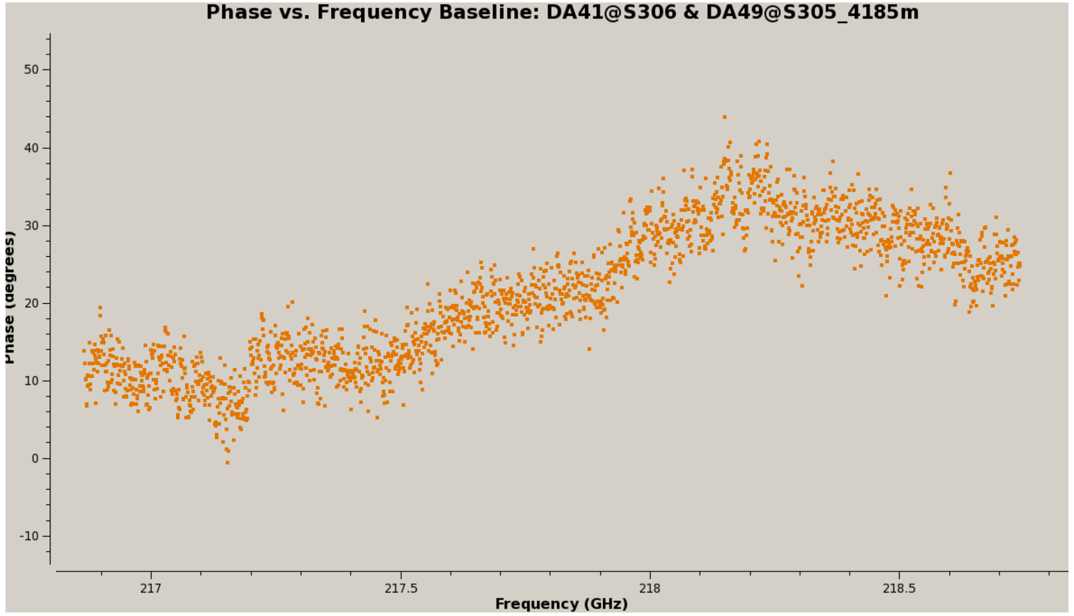
Part 1

Bandpass correction

- **Per antenna : Solve amplitudes and phases with frequency**



AMPLITUDE



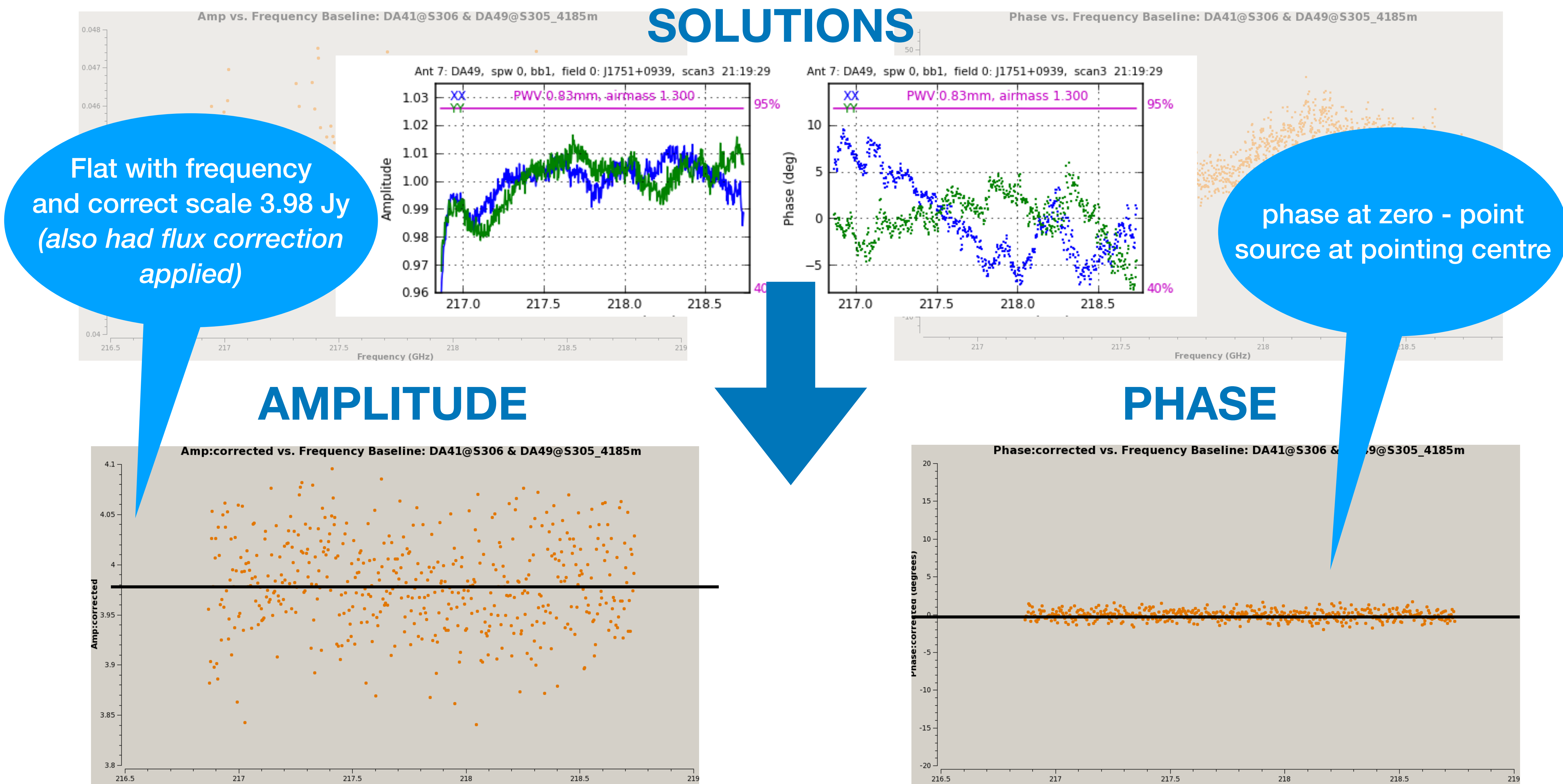
PHASE

take note of the scale

Part 1

Bandpass correction

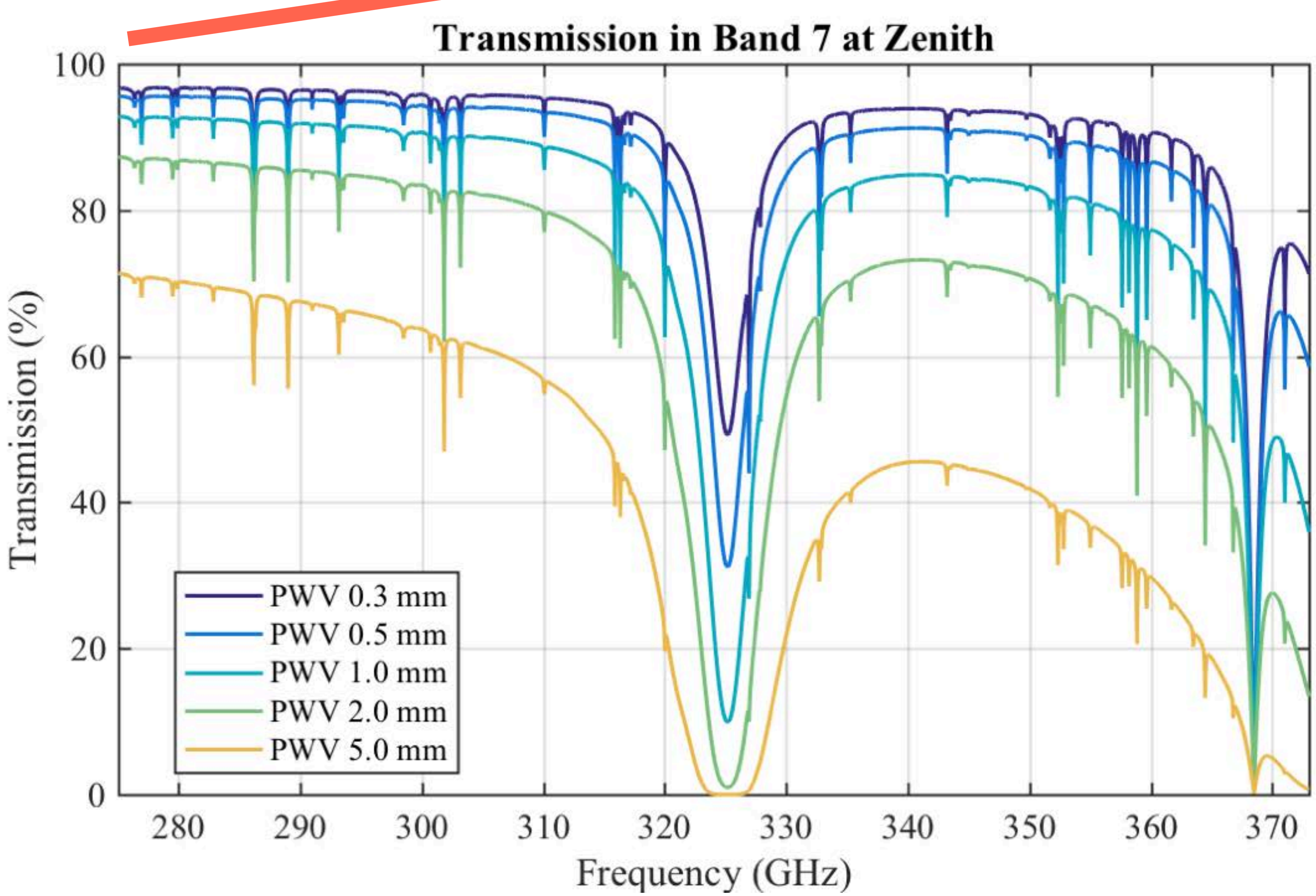
- **Per antenna : Solve amplitudes and phases with frequency**

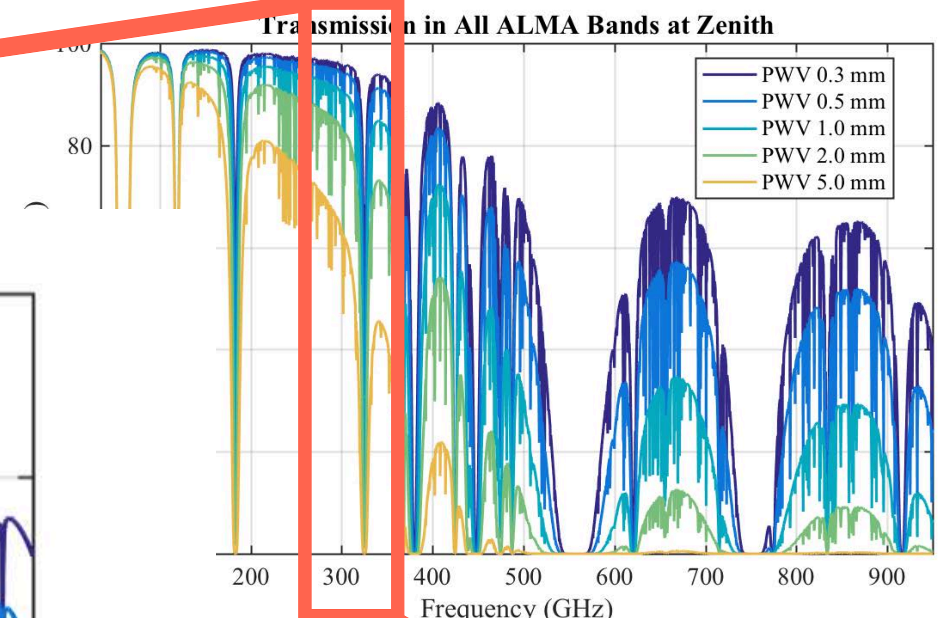


Part 1

- **System Temperature : Amplitude scale correction for receiver and sky**

  - correctly scale from instrument units to flux also accounting for looking through the atmosph

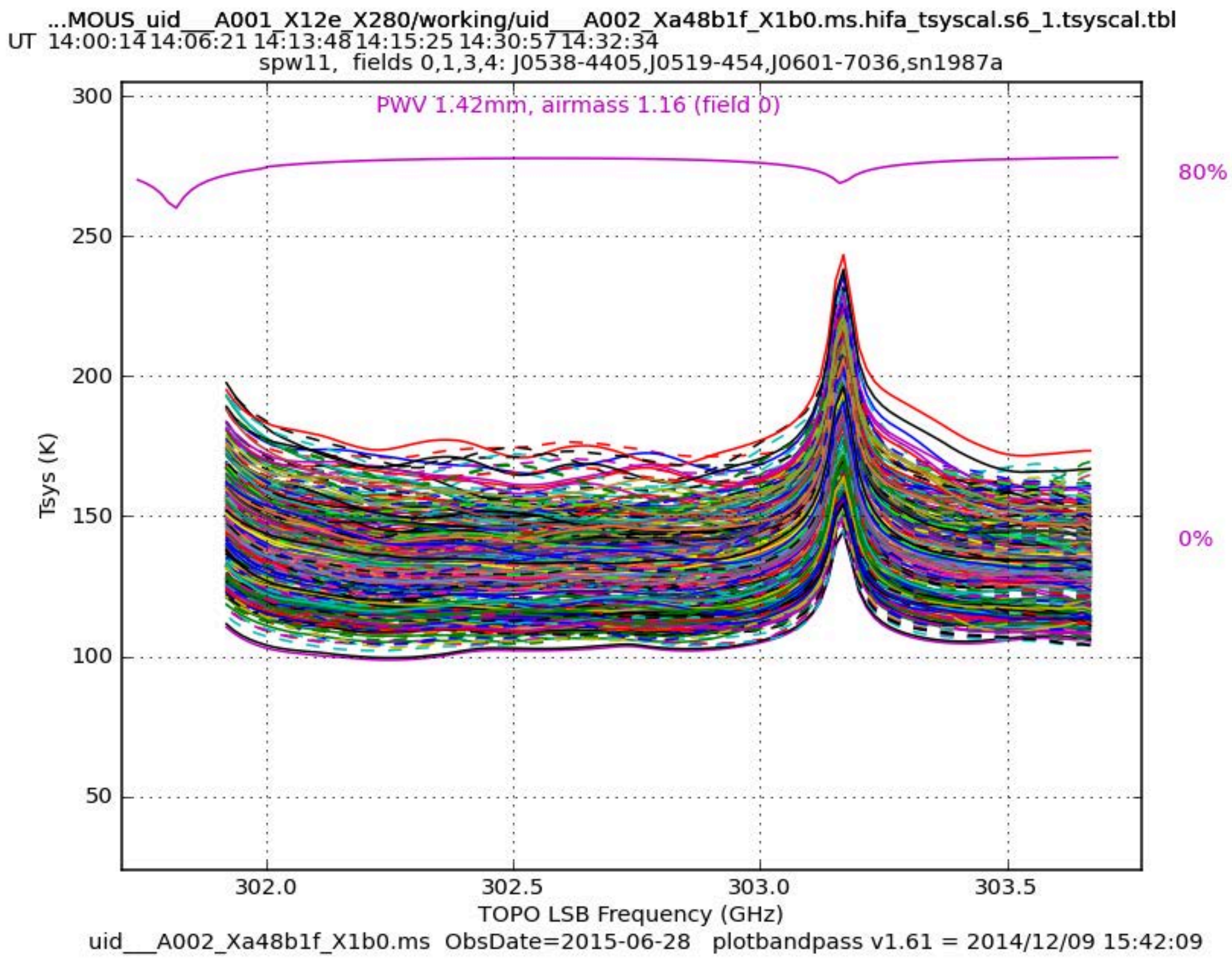




Part 1

- **System Temperature : Amplitude scale correction for receiver and sky**

  - correctly scale from instrument units to flux also accounting for looking through the atmosphere



per Antenna/Reciever per SpW

Scaling converts the ratio of **correlated signal** with the total system noise in a Kelvin (or Jansky) unit