Introduction to VLBI processing software $\mathcal{PIMA}$

Leonid Petrov *NASA GSFC*

Why *PIMA* ?

PIMA was developed for processing

1. survey style experiments (more than 10 sources)
2. processing data with spanned IFs
3. processing absolute astrometry/geodesy experiments

Gradually evolved for processing imaging experiments as well.

	Geodesy	Abs astrometry	Imaging	Diff. astrometry
AIPS	**rudimentary**	**incomplete**	**yes**	**yes**
HOPS	**yes**	**incomplete**	**no**	**no**
PIMA	**yes**	**yes**	**yes**	**incomplete**

Principles of $\mathcal{PIMA}$

- expects visibilities in FITS-IDI format
- batch-oriented
- modifiable
- written in a modern Fortran
- fast
- documented
- scriptable. Python wrappers are provided.

PIMA interface

Usage: pima <control_file> <operation> [options]

Control file consists of lines `keyword: value`

Example:

```
# PIMA_CONTROL file.  Format Version of 2016.10.19
#
# Created     on 2016.11.21_20:23:44
# Last update on 2016.11.21_20:23:56
#
SESS_CODE:      bp192j8
BAND:           X
#
UV_FITS:            /s0/vlba_fits/bp192j8/2016_09_07_bp192j8_01.fits
UV_FITS:            /s0/vlba_fits/bp192j8/2016_09_07_bp192j8_02.fits
UV_FITS:            /s0/vlba_fits/bp192j8/2016_09_07_bp192j8_03.fits
UV_FITS:            /s0/vlba_fits/bp192j8/2016_09_07_bp192j8_04.fits
#
STAGING_DIR:        NO
SOU_NAMES:          /vlbi/vcs9/vcs9_sou.names
STA_NAMES:          /vlbi/solve/save_files/vlbi_station.names
PCAL:               USE_ALL
TSYS:               CLEANED
GAIN:               USE
SAMPLER_CAL:        USE
#
WARNING:            YES
DEBUG_LEVEL:        2

...
```

$\mathcal{PIMA}$ main tasks

- load — load the data
- gean — GEt ANtenna calibration
- frib — FRInge fitting, baseline mode
- bpas — Generation of complex bandpass
- mkdb — MaKe DataBase
- splt — split the time- and frequency-averaged visibilities

PIMA task load

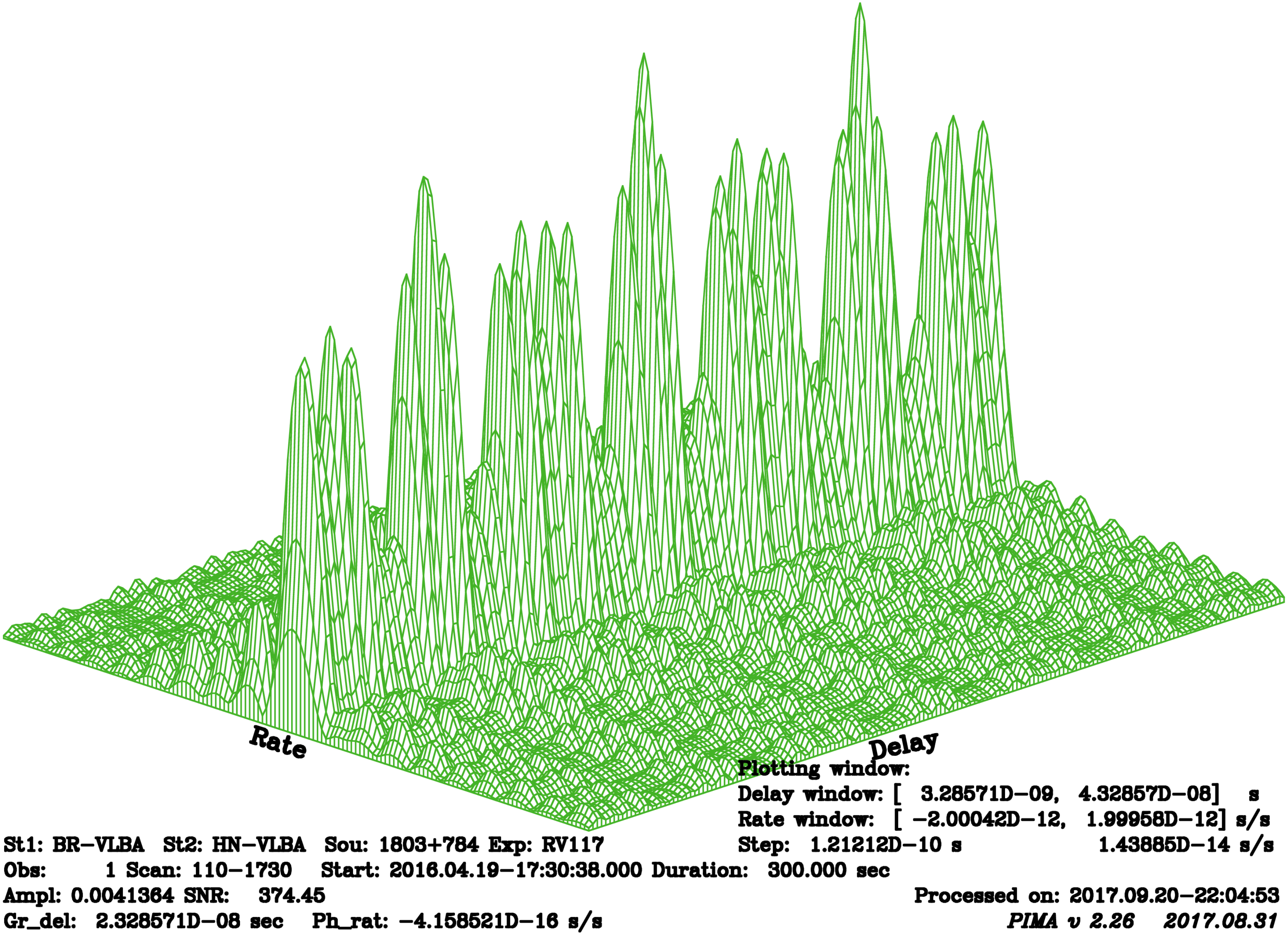
- *PIMA* does not re-write visibility data in its own format
- *PIMA* creates numerous indexing tables and use them for reading visibility data
- Task load creates a number of files describing data contents and statistics
- Task load checks data for consistency
- Task load splits the data into scans
- Task load renames, splits and/or merges sources
- Task computes its own a priori model using VTD

$\mathcal{PIMA}$ task gean

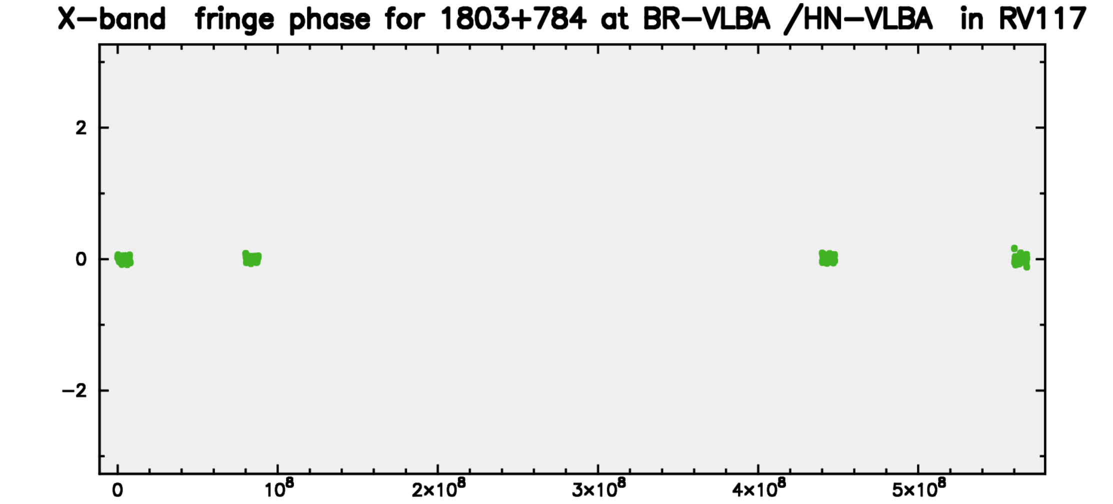
- Parses log files and creates its own antab flavor
- Loads calibration information
- Prints calibration information

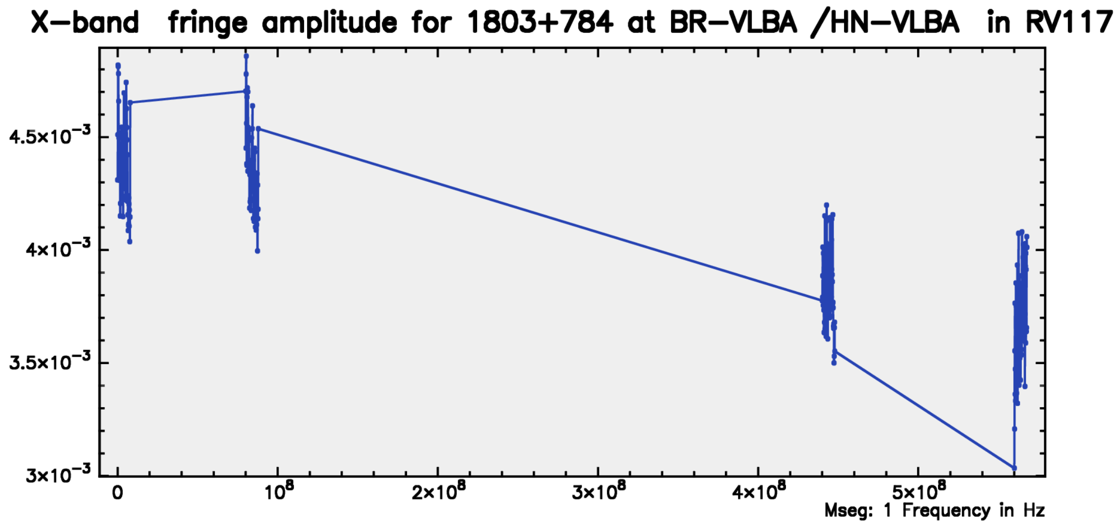
$\mathcal{PIMA}$ task frib — main horse

- runs fringe fitting in baseline mode
- performs coarse fringe fitting with a single 2D FFT without prior computation of single-band delay
- performs fine LSQ fringe fitting with additive and multiplicative re-weighting
- computes group delay rate or phase acceleration
- supports over-sampling
- computes noise statistics
- supports I-polarization data on the fly
- supports a priori phase rotation for sources with large a priori errors
- generates fringe plots
- generates output ascii file
- supports OBS: ALL, OBS: obs_ind, OBS: range, INCLUDE_OBS_FILE, EXCLUDE_OBS_FILE

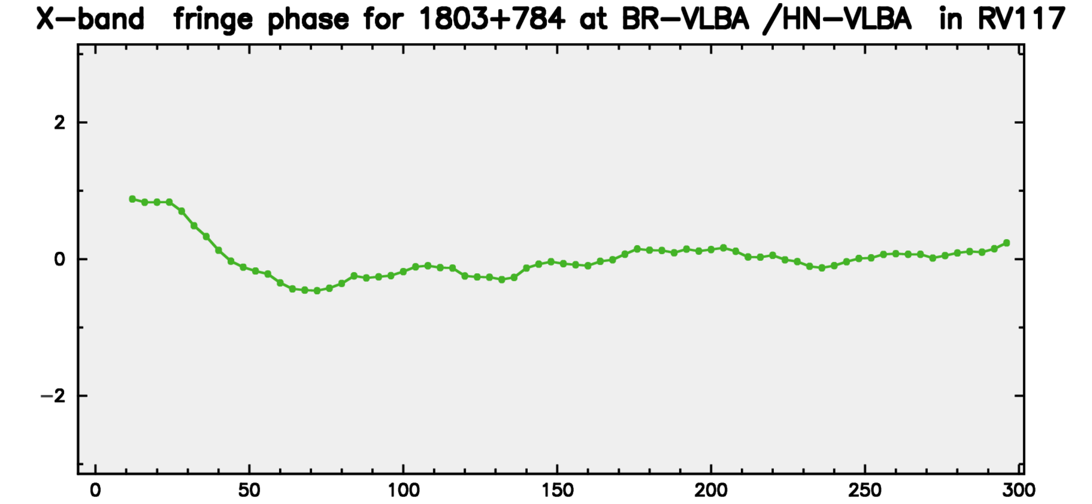

Rate
Delay
Plotting window:
Delay window: [3.28571D-09, 4.32857D-08] s
Rate window: [-2.00042D-12, 1.99958D-12] s/s
Step: 1.21212D-10 s 1.43885D-14 s/s
St1: BR-VLBA St2: HN-VLBA Sou: 1803+784 Exp: RV117
Obs: 1 Scan: 110-1730 Start: 2016.04.19-17:30:38.000 Duration: 300.000 sec
Ampl: 0.0041364 SNR: 374.45
Processed on: 2017.09.20-22:04:53
Gr_del: 2.328571D-08 sec Ph_rat: -4.158521D-16 s/s
PIMA v 2.26 2017.08.31

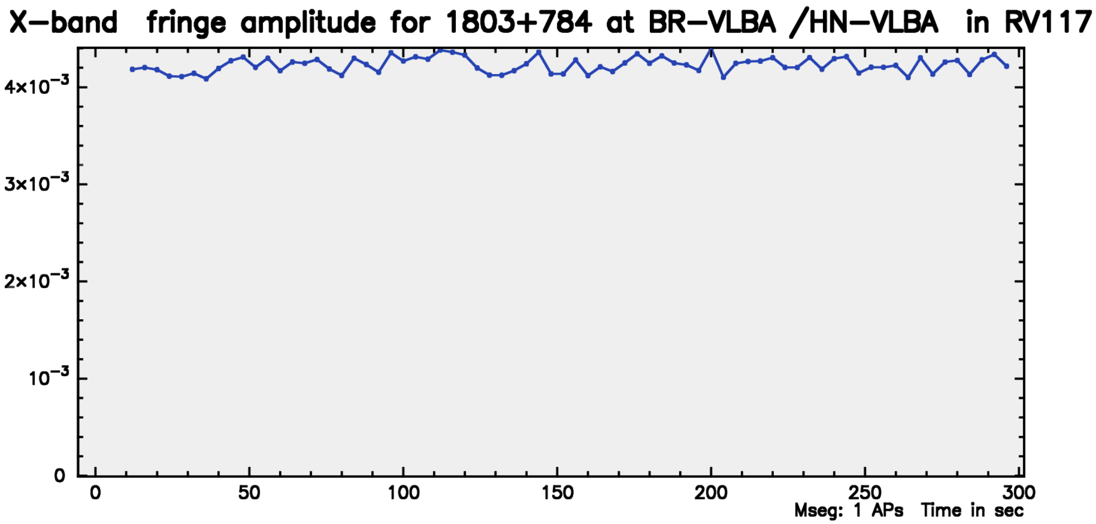
Fringe phs/amp for obs #1 Exp RV117 SNR=374.4





Fringe phs/amp for obs #1 Exp RV117 SNR=374.4





$\mathcal{PIMA}$ task bpas

- Allows data visualization
- Usually the task runs in a conjunction with bandpass mask computation of single-band delay
- Runs in three modes: init, accum, fine.
- Supports polarization bandpass

$\mathcal{PIMA}$ task mkdb

- computes scan reference time
- computes total group delay, phase delay, phase delay rate
- supports output in vgosda (GVF) format (interface with Solve) and in plain ascii
- Implements interface with Solve

$\mathcal{PIMA}$ task splt

- applies numerous calibrations and normalizations
- applies phase rotation according to fringe fitting results
- splits data source by source
- applies specifies time and frequency averaging
- writes down in fits format
- Implements interface with Difmap

PIMA data processing pipeline

- Creation of control file, check of catalogue files
- Creation of indexing tables with *PIMA* task load
- Examination of dumps and statistics
- Parsing log files and loading them in *PIMA*
- Phase cal tone masking
- Coarse fringe fitting with *PIMA* task frib
- Bandpass calibration with *PIMA* task bpas
- Fine fringe fitting with *PIMA* task frib
- Creation of the output database file with *PIMA* task mkdb
- Running VTD/Post-Solve. Flagging data and interactive astro/geo solution
- Generating control file for re-fringing
- Re-fringing affected data
- Creation of the output database file with *PIMA* task mkdb
- Running VTD/Post-Solve. Flagging data and interactive astro/geo solution

- Generating control file for computation of the atmosphere opacity and Tatm with task opag
- Loading atmosphere opacity and Tatm with task opal
- Tsys cleaning with task tsmo
- Detection antenna off-source running $\mathcal{PIMA}$ task onof
- Identifying 2–4 reference sources
- Running $\mathcal{PIMA}$ task splt for the reference sources
- Imaging reference sources with Difmap
- Computation of gain correction with task gaco
- Running $\mathcal{PIMA}$ task splt for remaining sources
- Imaging remaining sources with Difmap in automating mode
- Examining images and re-imaging them with Difmap manually

Theoretical basis

Received emission:

$$X_1 = S + n_1$$

$$X_2 = S\, e^{-i\phi(\omega,t)} + n_2$$

Typically, $\dfrac{|S|}{|n_i|} = 10^{-4}$ – 10^{-3}

Maximum likelihood estimation:

$$L(X_1, X_2 | \phi(\omega, t)) \longrightarrow \max$$

n_i and S are Gaussian complex processes.

Assuming $\mathrm{Cov}(n_i, S) = 0$ and $\mathrm{Cov}(n_{ij}, n_{ik}) = 0$

This results to

$$\sum_k X_1\, X_2^*\, e^{i\phi(\omega,t)} \longrightarrow \max$$

Let us have cross-correlation spectrum $c_{kj}(t,\omega)$ with weights w_{kj}. We search for those $\tau_p, \tau_g, \dot{\tau}_p, \dot{\tau}_g)$ that maximize the spectrum averaged over time and frequency:

$$C(\tau_p, \tau_g, \dot{\tau}_p, \dot{\tau}_g) =$$

$$\sum_k \sum_j c_{kj}(t,\omega)\, w_{kj} \times e^{i(\omega_0 \tau_p \,+\, \omega_0 \dot{\tau}_p (t_k - t_0) \,+\, (\omega_j - \omega_0)\tau_g \,+\, (\omega_j - \omega_0)\dot{\tau}_g (t_k - t_0))} =$$

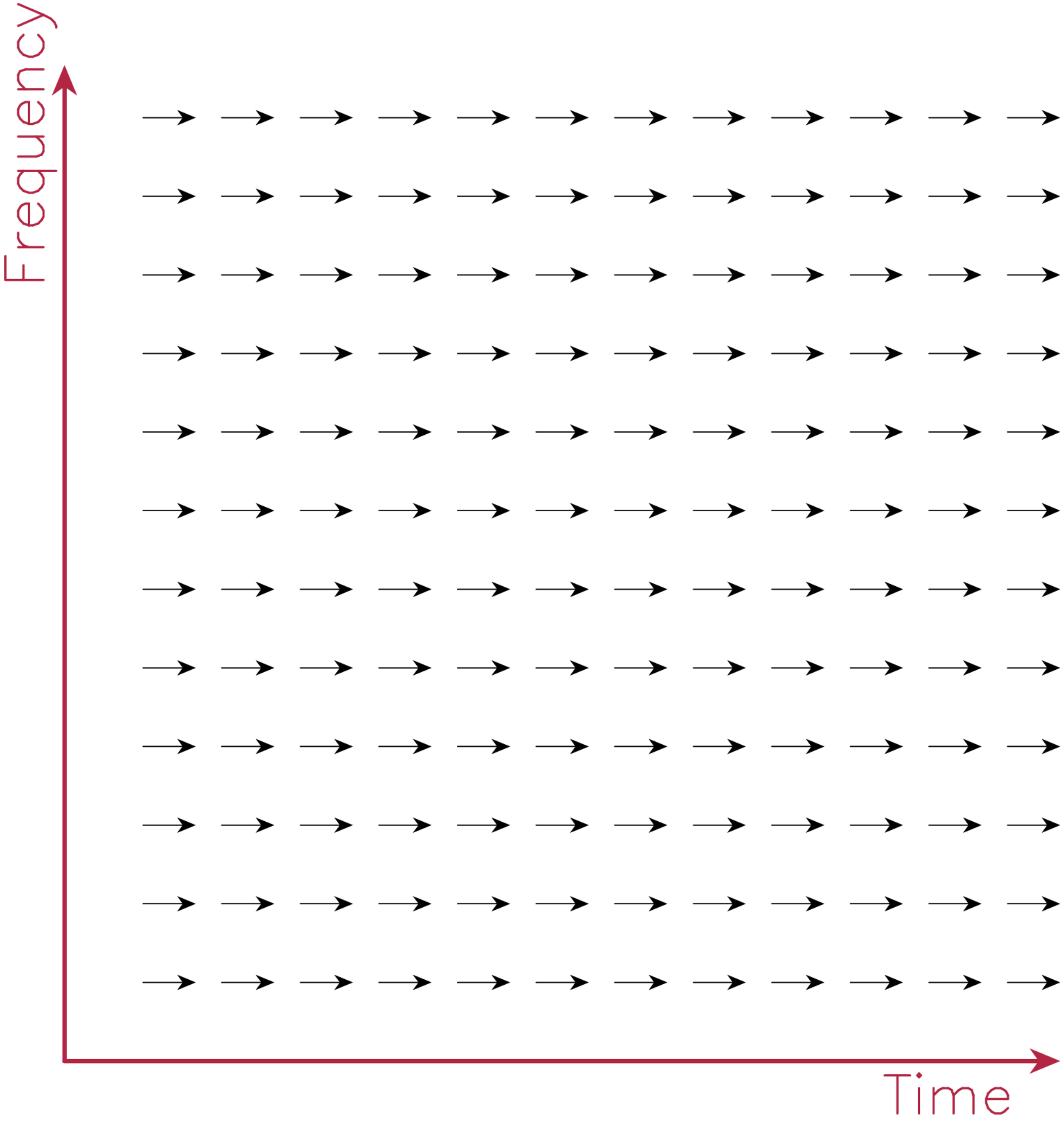
$$e^{i\omega_0 \tau_p} \sum_k \sum_j c_{kj}(t,\omega)\, w_{kj} \times e^{i(\omega_0 \dot{\tau}_p (t_k - t_0) \,+\, (\omega_j - \omega_0)\tau_g \,+\, (\omega_j - \omega_0)\dot{\tau}_g (t_k - t_0))}$$

Variants:

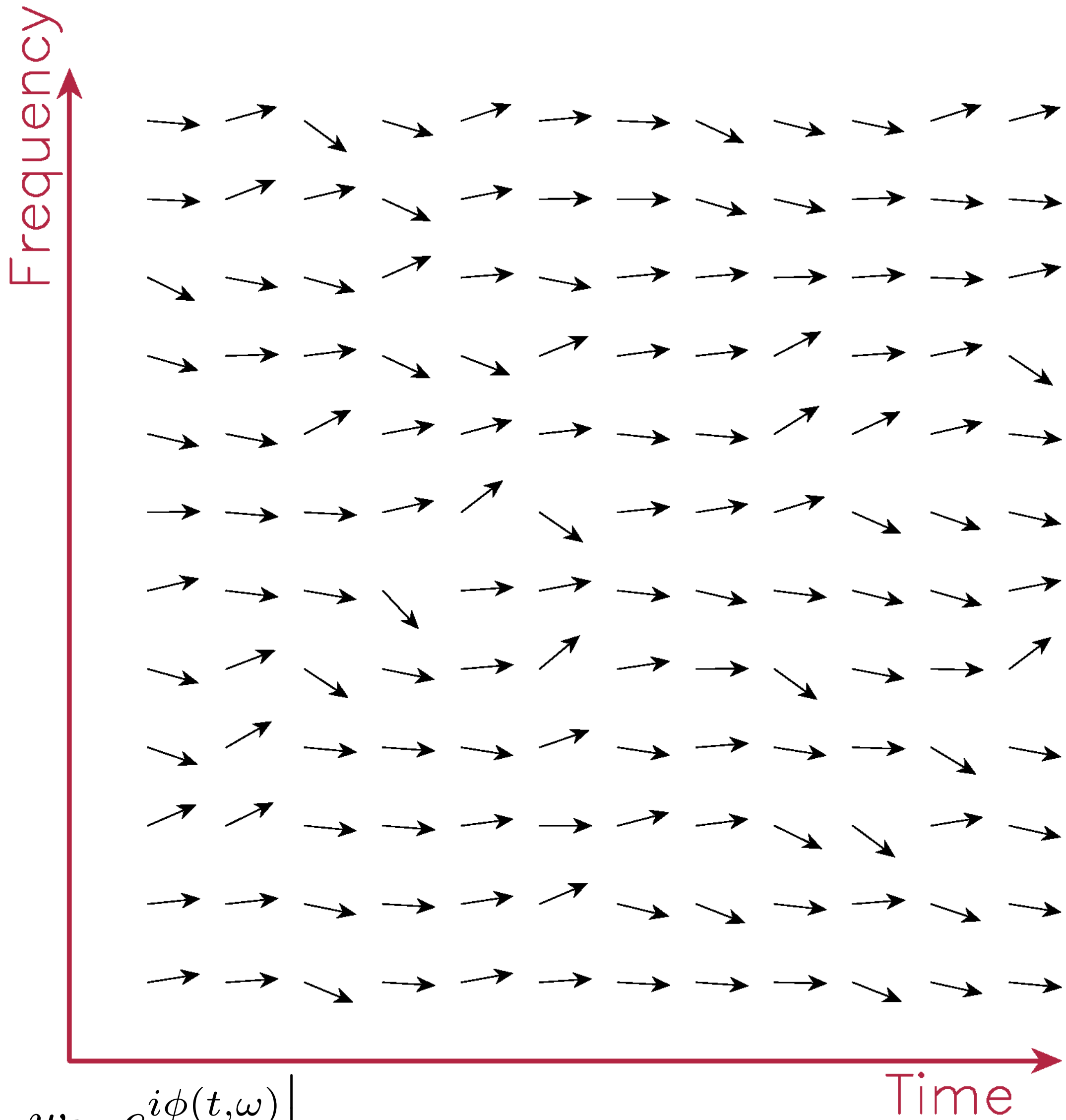
$$\phi'(t,\omega) = \omega_0 \dot{\tau}_p (t_k - t_0) \;+\; (\omega_j - \omega_0)\tau_g \;+\; \ddot{\tau}_p (t_k - t_0)^2/2$$

$$\phi'(t,\omega) = \omega_0 \dot{\tau}_p (t_k - t_0) \;+\; (\omega_j - \omega_0)\tau_g \;+\; \frac{1}{\omega_j}\,\epsilon$$

Time-frequency diagram. Ideal case:

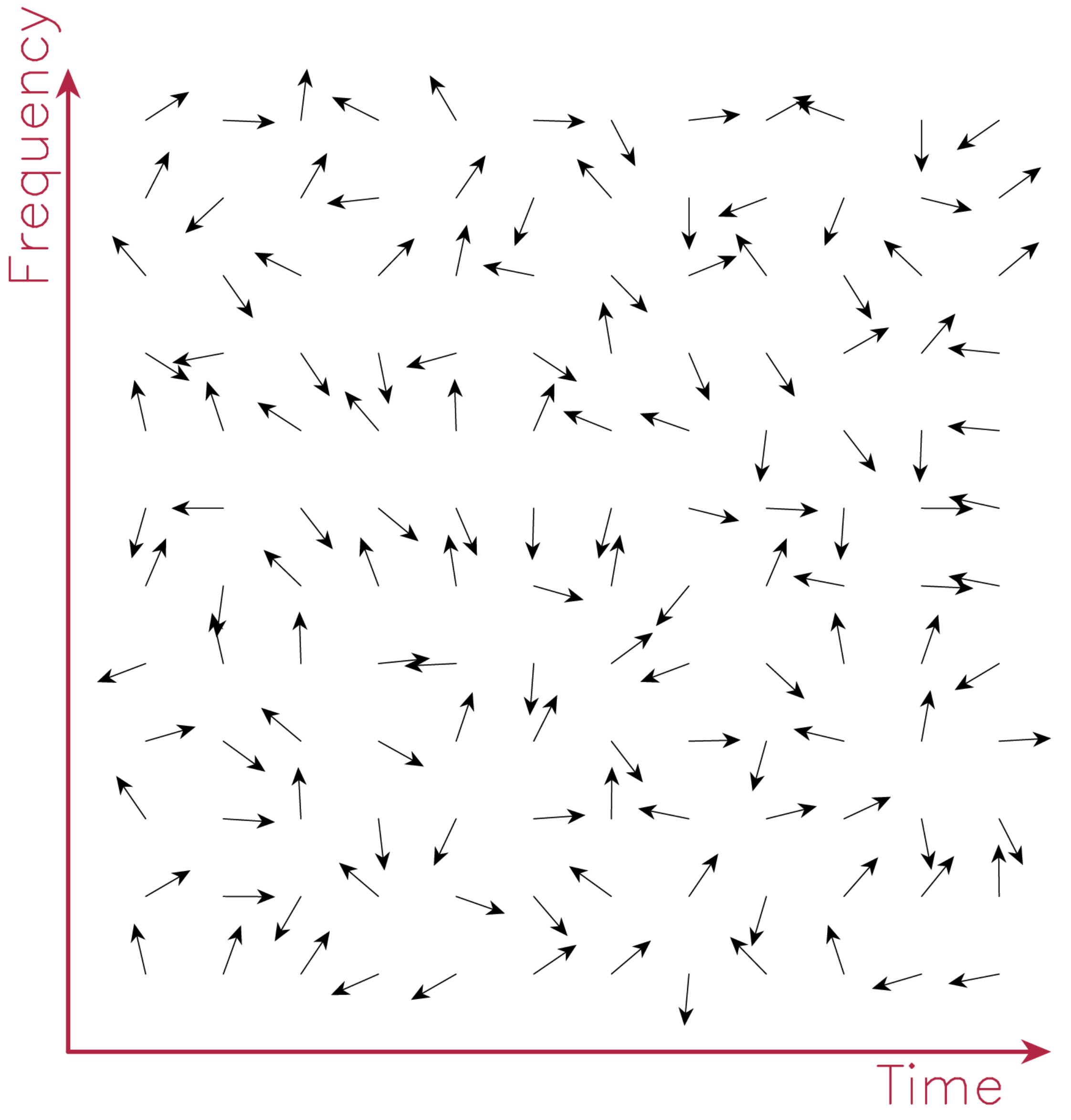


Time-frequency diagram. Weak phase noise is added

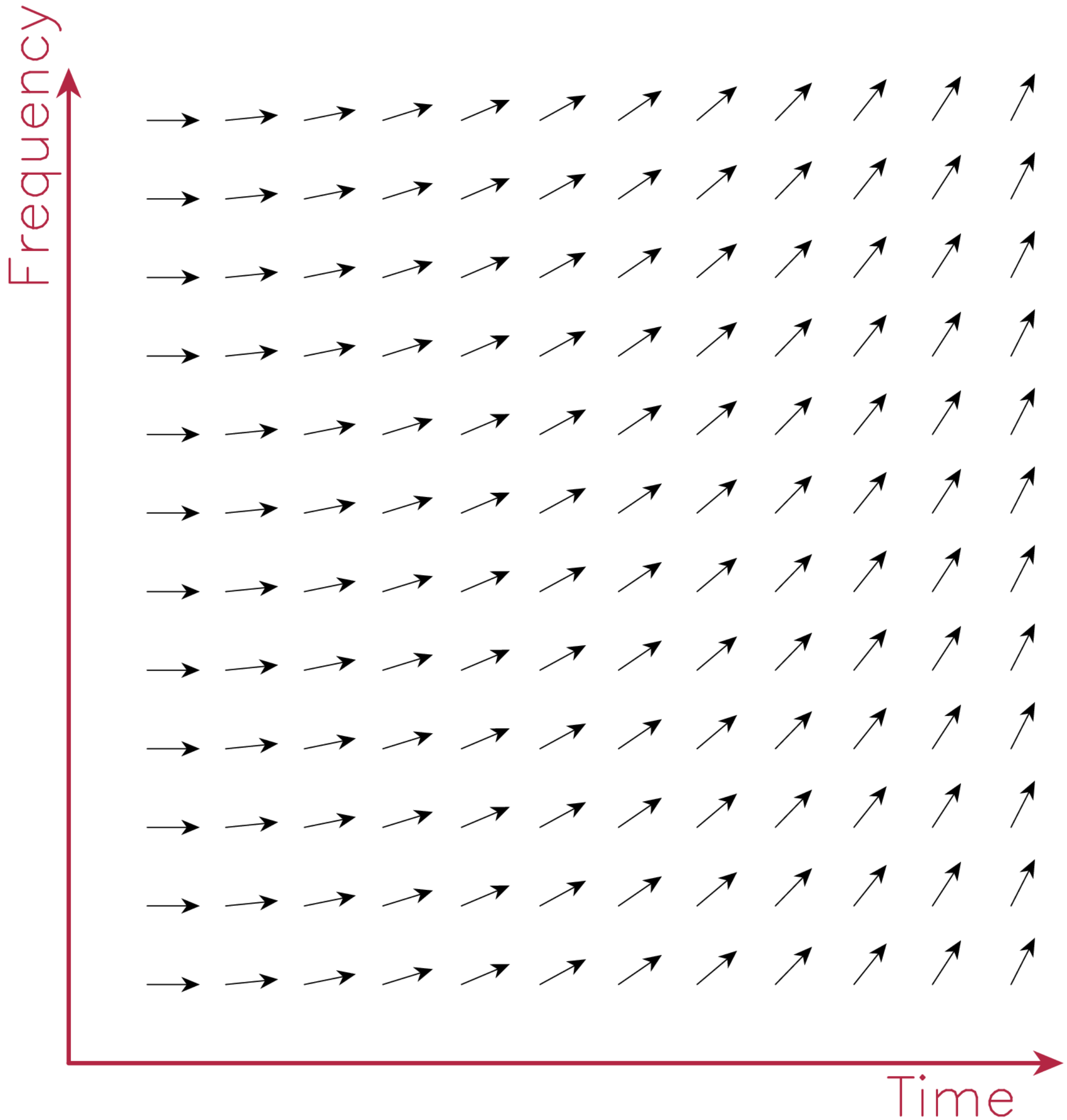


$$\text{Decor} = \frac{\left|\sum c_{ij} w_{kj}\, e^{i\phi(t,\omega)}\right|}{\sum |c_{kj}\, w_{kj}|}$$

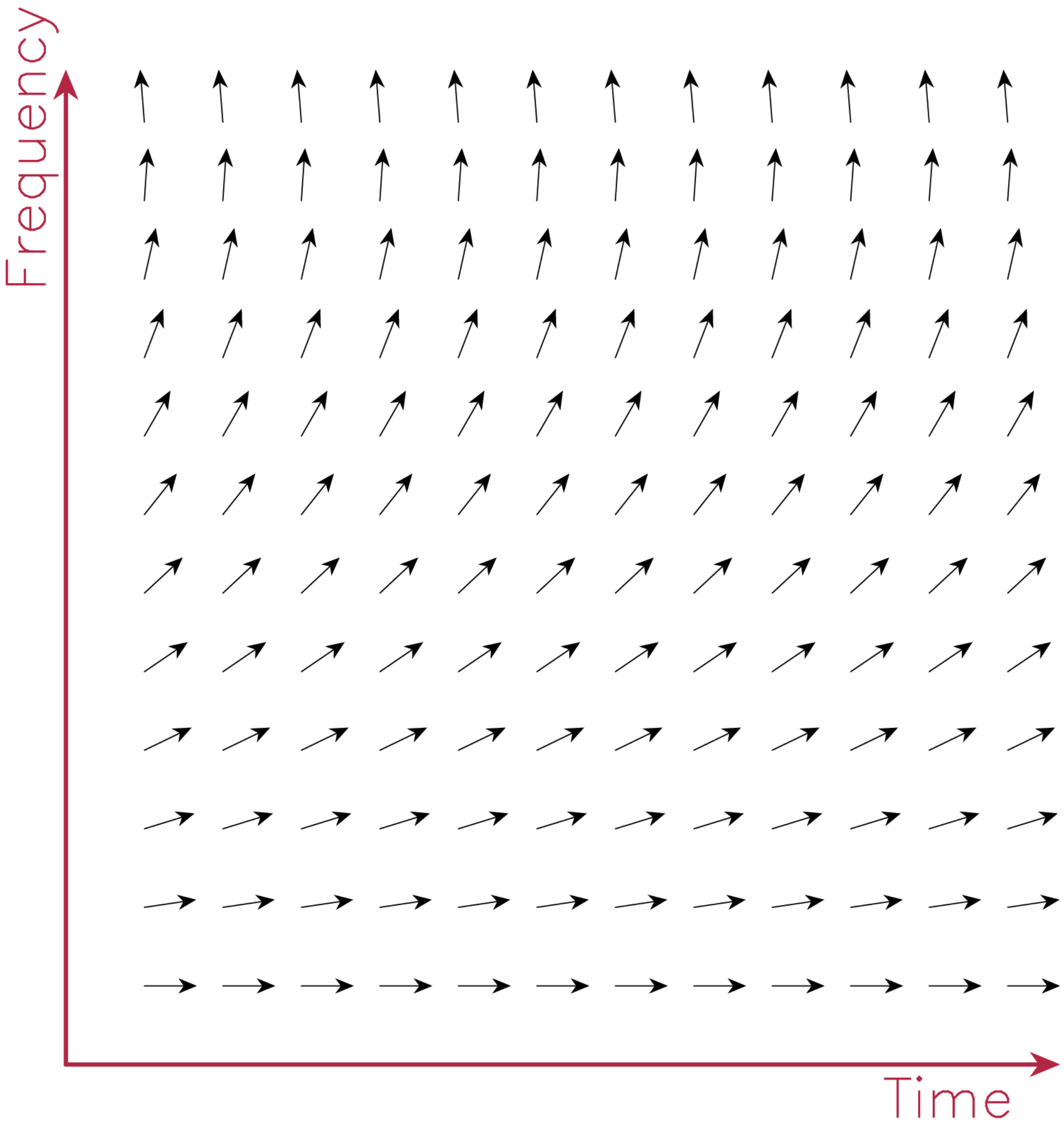
Time-frequency diagram. Strong phase noise is added



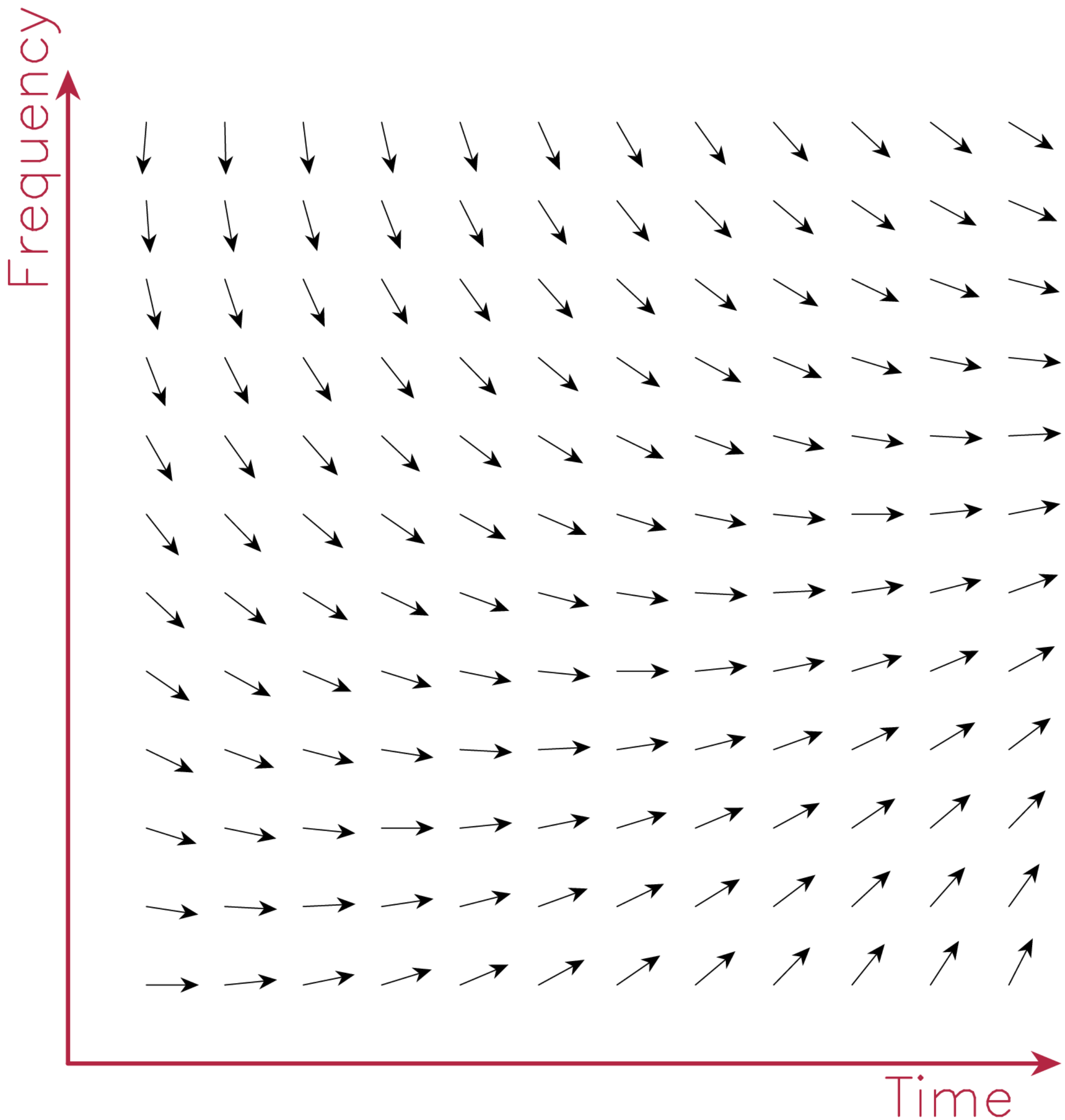
Time-frequency diagram. Phase rate $\neq 0$. No noise is added.



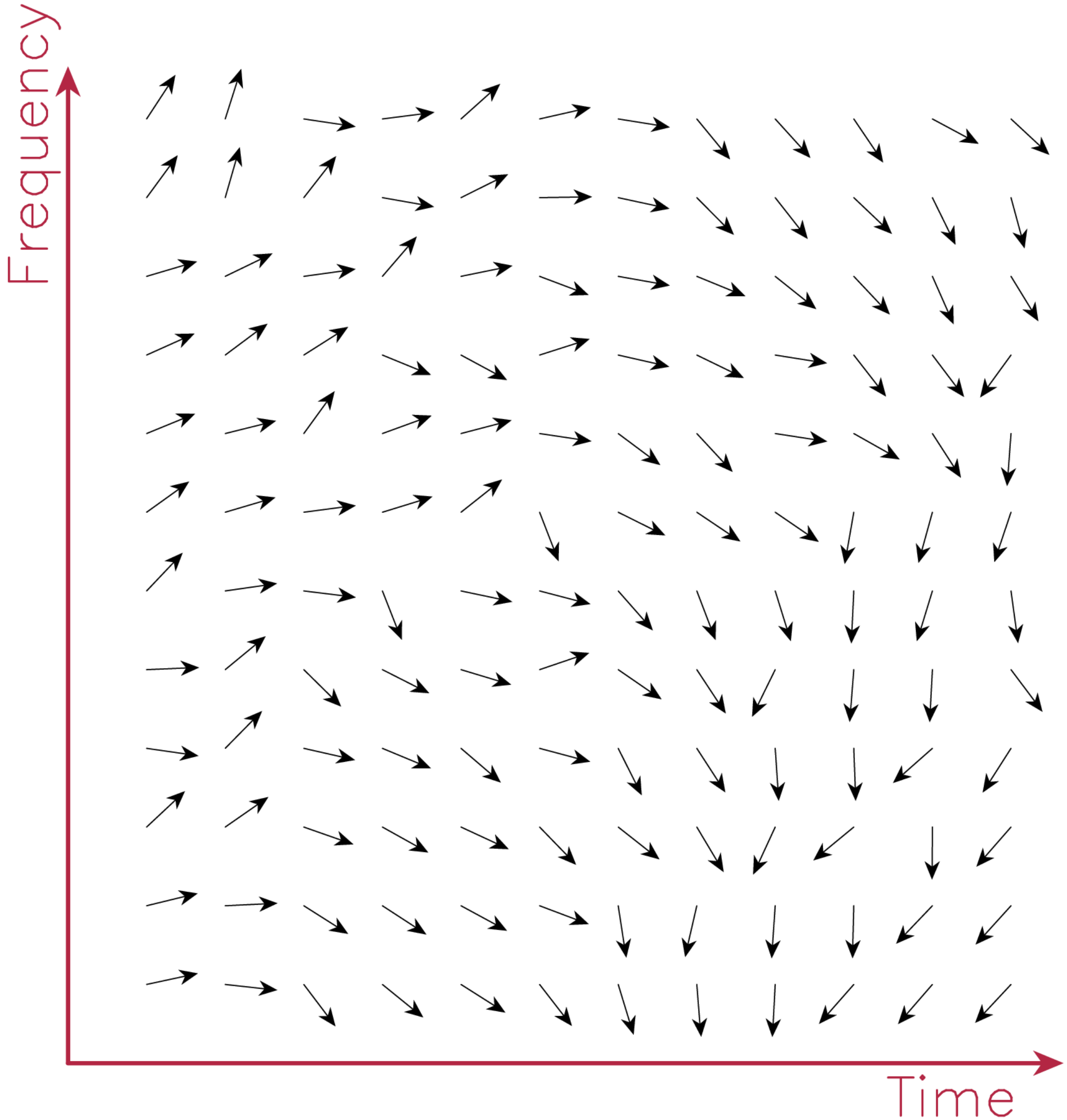
Time-frequency diagram. Group delay $\neq 0$. No noise is added.



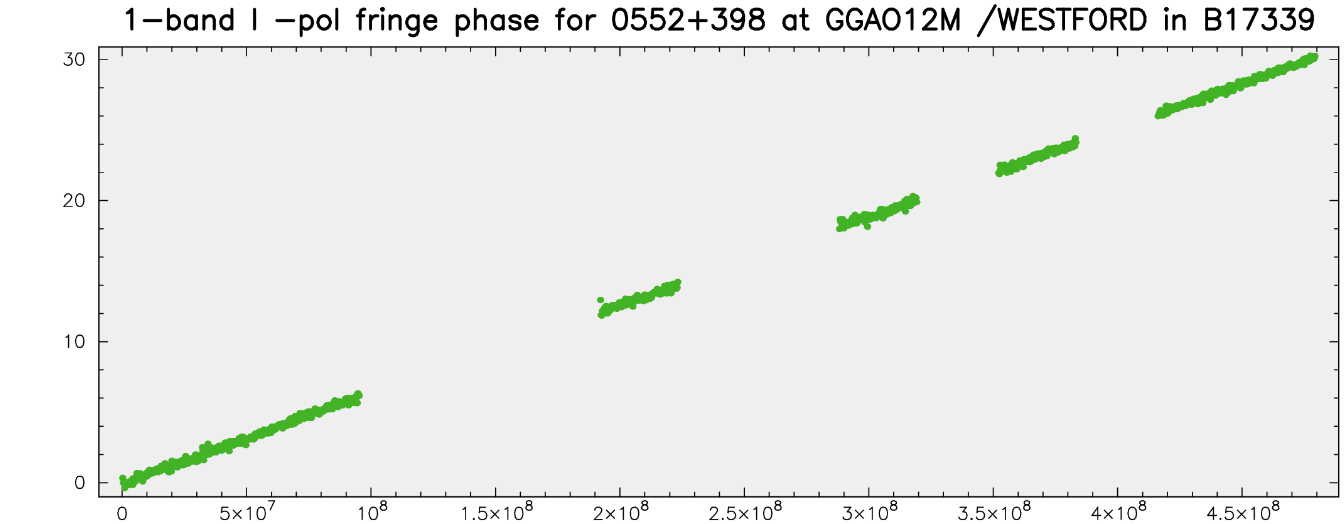
Time-frequency diagram. Group delay $\neq 0$. Phase delay rate $\neq 0$. No noise is added.

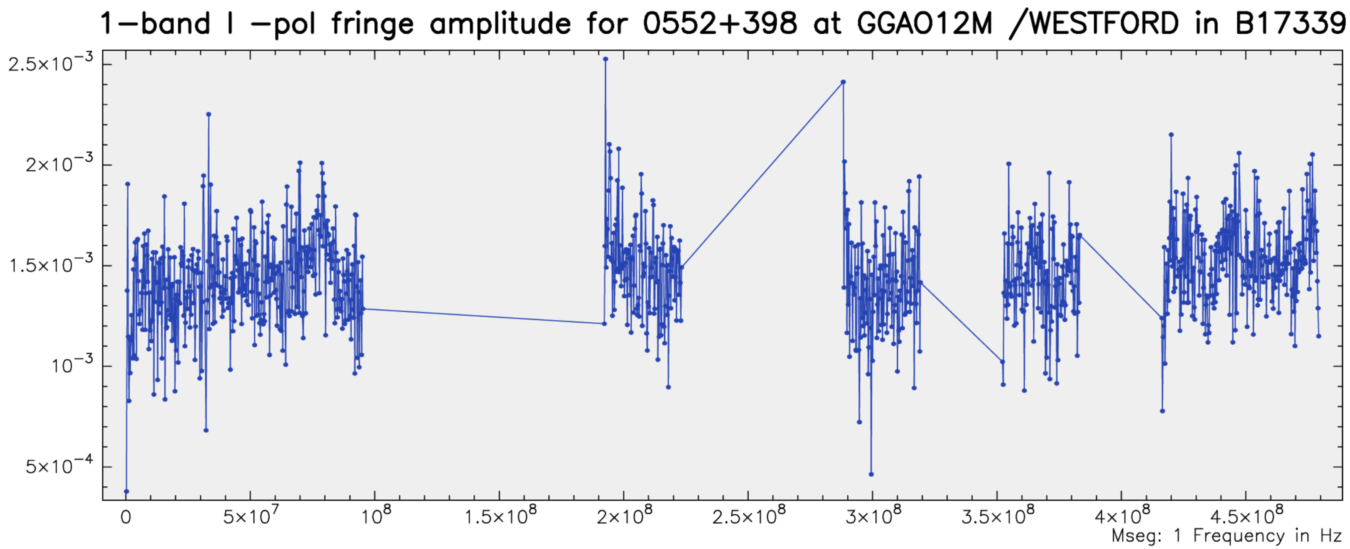


Time-frequency diagram. Group delay $\neq 0$. Phase delay rate $\neq 0$. Noise is added.

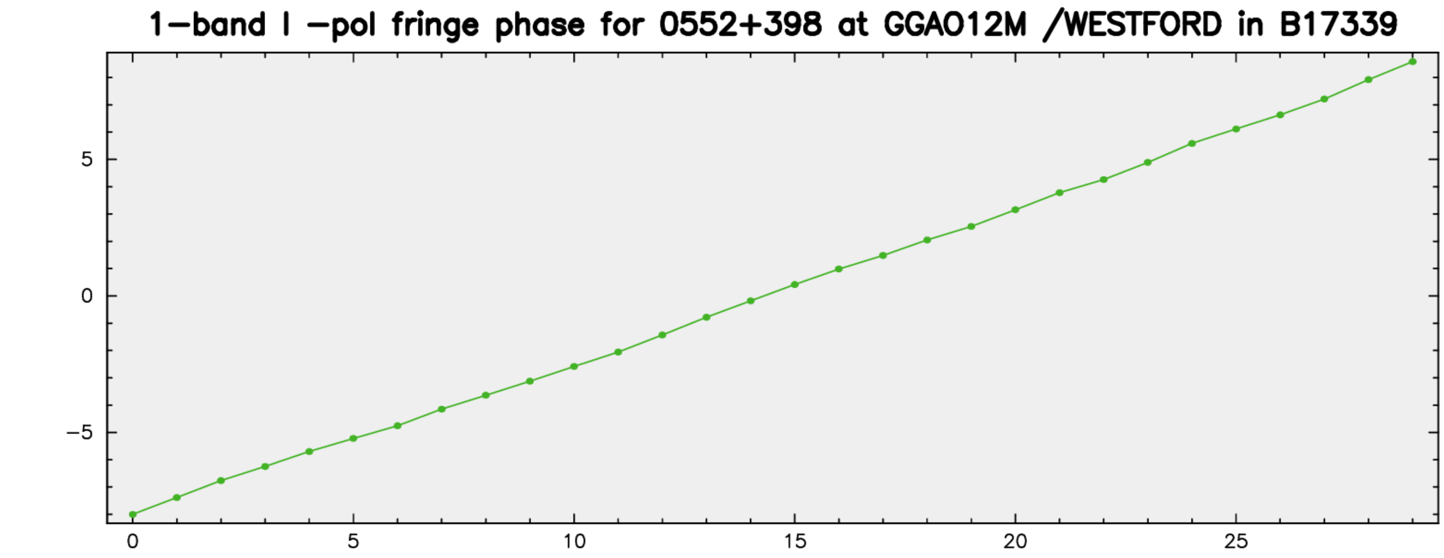


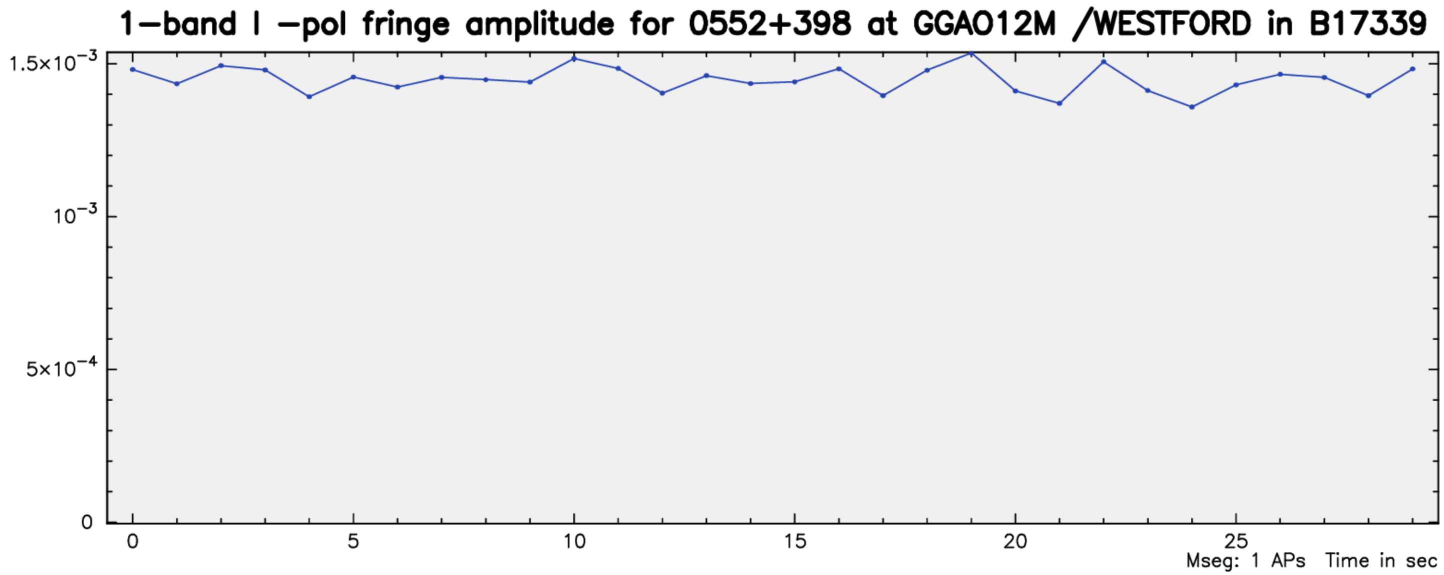
Real data. Phase/amplitude spectrum





Real data. Time-dependent phase and amplitude





Coarse fringe fitting

Coarse fringe search through 2D or 3D parameter space

2D parameter space, $\tau_g, \dot{\tau}_p$:

$$C'(\dot{\tau}_p, \tau_g, \dot{\tau}_g, \ddot{\tau}_p, \epsilon)\, e^{-i\omega_0\tau_p} = \sum_k \sum_j c_{kj}(t,\omega)\, w_{kj} \times e^{i(\omega_0\dot{\tau}_p(t_k - t_0) \,+\, (\omega_j - \omega_0)\tau_g)}$$

$$= \mathcal{F}(c\, w)$$

The 3rd parameter, $\dot{\tau}_g, \ddot{\tau}_p, \epsilon$ is estimated on the outer cycle on a discrete grid.

Oversampling

Sampling intervals are $\Delta t/\beta$ and $\Delta f/\gamma$ where Δt and Δf

$$L = \frac{1}{t_s}\int_{-t_s/2}^{t_s/2} \cos 2\pi\left(\omega_0\tau_p - \frac{k}{\beta t_s}\right) t\, dt \;\times\; \frac{1}{f_b}\int_{-f_b/2}^{f_b/2} \cos 2\pi\left(\omega_0\tau_g - \frac{j}{\gamma f_b}\right) f\, df =$$

$$= \operatorname{sinc}\left(\pi/\left(2\beta\right)\right) \cdot \operatorname{sinc}\left(\pi/\left(2\gamma\right)\right)$$

$L_{\min} = \frac{4}{\pi^2} = 0.405$ when $\beta = 1$, $\gamma = 1$

$L_{\min} = 0.949$ when $\beta = 4$, $\gamma = 4$

Fine fringe fitting

Least squares in the vicinity of the coarse fringe fitting solution

- Run coarse solution
  1. Variant
     - Outer cycle over the 3rd parameter
     - 2D FFT
     - Select the 3rd parameter that maximizes $\sum_k \sum_j c_{kj}(t,\omega)\, w_{kj}\, e^{i\phi(t,\omega)}$
  2. Variant
     - 2D FFT
     - Inner cycle over the 3rd parameter
     - Select the 3rd parameter that maximizes $\sum_k \sum_j c_{kj}(t,\omega)\, w_{kj}\, e^{i\phi(t,\omega)}$

- Apply results of coarse fringe

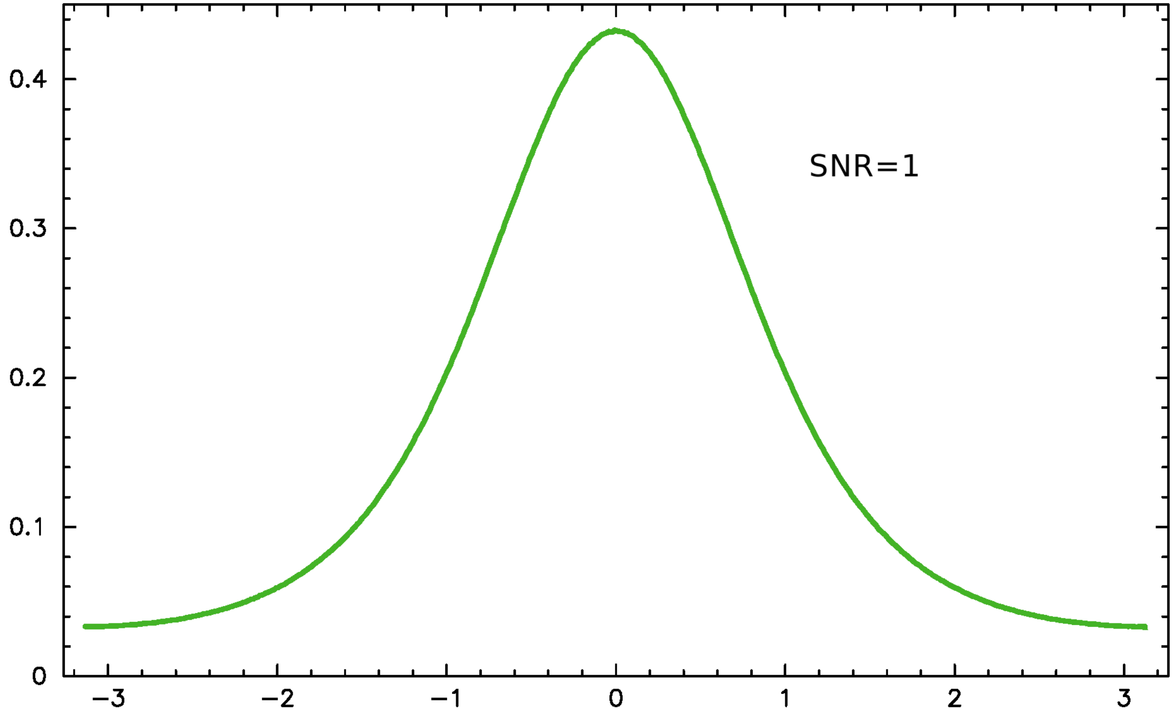
- Get residual phases and estimate corrections for 4-parameters: $\tau_{ph}, \tau_{gr}, \dot{\tau}_{ph}$, $\dot{\tau}_{gr}$, or $\ddot{\tau}_{ph}$, or ϵ using either a parabolic fit or LSQ.

  Weights: $|c_{kj}\, w_{kj}|$.

Reweighing

Variance of fringe phase

When SNR $\ll 1$, the distribution is uniform.



When SNR $\gg 1$, the distribution is Gaussian and $\sigma(\phi) = \sqrt{\frac{2}{\pi}} \frac{1}{\mathrm{SNR}}$

under assumption that visibilities are independent

$\mathcal{PIMA}$ fine fringe search computes

1. fringe phase
2. phase delay rate
3. group delay
4. one of
   - group delay rate
   - phase acceleration
   - Δ TEC

Before fine search, $\mathcal{PIMA}$ applies results of the coarse search and pre-averages the data over time and frequency into segments. Segment size is chosen to have SNR≈ 1.

Three variants of using segment weights

- weights are as is $w_{ij} = \sqrt{\frac{\pi}{2}}\,\mathrm{SNR}$
- multiplicative re-weighting $\alpha\, w_{ij}$
- additive re-weighting $\sqrt{\alpha^2 + w_{ij}^2}$

Phase calibration

$\mathcal{PIMA}$ supports 1 tone or multiple tones

- Single pcal tone per IF: the same pcal phase for entire IF is assumed
- Case of multiple tones: tones affected by spurious signals are masked out.
- Pcal phase is interpolated/extrapolated
- Pcal can be enabled/disabled for given station(s)

Amplitude (re)-normalization

Correction for digitization:

$$
\begin{aligned}
\rho_{out} \quad = \quad & 2\,\kappa \int\limits_0^{\rho} \frac{1}{\sqrt{1-\rho^2}} d\rho \\
& + 2\,\kappa\,(n-1) \int\limits_0^{\rho} \frac{1}{\sqrt{1-\rho^2}} \left(e^{-\frac{v^2}{2(1-\rho^2)}} + e^{-\frac{v_1^2}{2(1-\rho^2)}} \right) d\rho \\
& + \kappa\,(n-1)^2 \int\limits_0^{\rho} \frac{1}{\sqrt{1-\rho^2}} \left(e^{-\frac{v_1^2 - 2\rho v_1 v_2 + v_2^2}{2(1-\rho^2)}} + e^{-\frac{v_1^2 + 2\rho v_1 v_2 + v_2^2}{2(1-\rho^2)}} \right) d\rho,
\end{aligned}
$$

Table 1: Numerical coefficients in integral 5 for three cases of the number of bits per sample: (1,1), (1,2), (2,2)

	n	v_1	v_2	κ
(1,1)	1.0	0.0	0.0	0.3803
(1,2)	3.3359	0.0	0.9816	0.05415
(2,2)	3.3359	0.9816	0.9816	0.07394

- Correct auto-correlation for digital distortion
  - inverse Fourier-transform to lag domain
  - de-tapering
  - extrapolation of to zero
  - digital correction
  - tapering
  - direct Fourier-transform to frequency domain
- Normalize autocorrelation: $N_i = \int A_i(\omega)\, d\omega$
- Apply autocorrelation normalized coefficients to cross-spectrum:

$$\frac{1}{\sqrt{N_1\, N_2}}$$

- Correct for cross-spectrum amplitude for digital distortion
- Re-normalize cross-spectrum amplitude for masking out channels.

$$\alpha = \frac{\int A(\omega)\, d\omega}{\int A(\omega)\, w(\omega)\, d\omega}$$

The probability of false association

In the absence of signal, the amplitude of the cross correlation function a has the Rayleigh distribution:

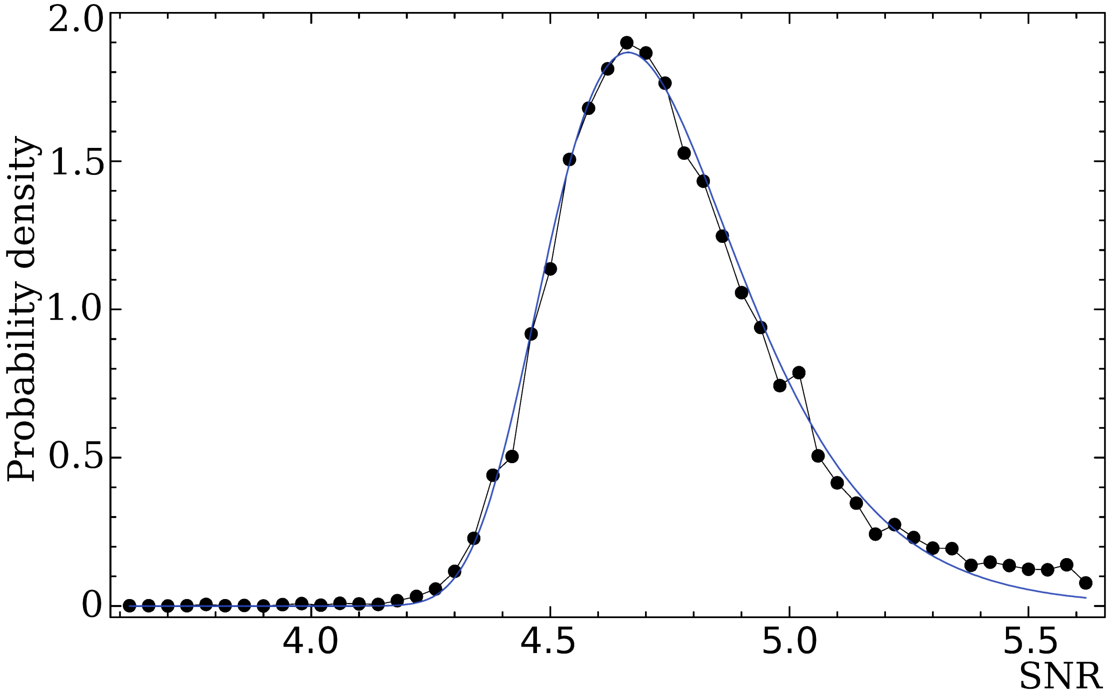
$$p(a) = \frac{a}{\sigma^2} e^{-\frac{a^2}{2\sigma^2}},$$

the cumulative distribution function of the coherent sum over n **independent** points is

$$P(a) = \left(1 - e^{-\left(\frac{a^2}{2\sigma^2}\right)}\right)^n.$$

Differentiating $P(a)$ over a, we get the SNR distribution in the absence of signal (s=a/σ):

$$p(s) = \sqrt{\frac{2}{\pi}} \frac{n}{\sigma} s \, e^{-\frac{s^2}{\pi}} \left(1 - e^{-\frac{s^2}{\pi}}\right)^{n-1}.$$



Fit n_{eff} and σ_{eff} to the empirical distribution

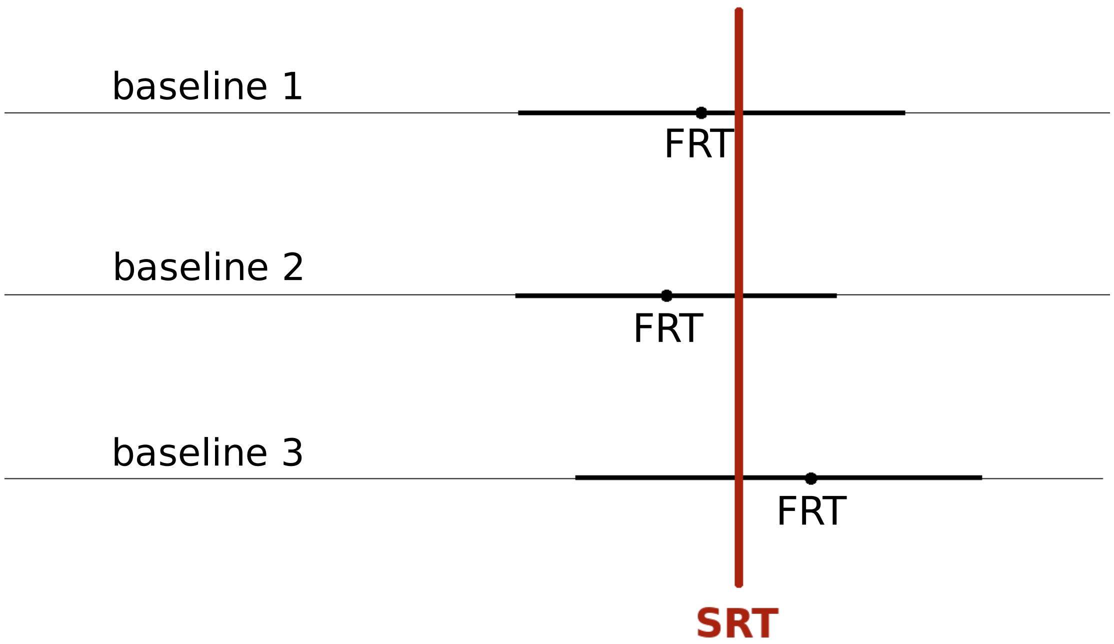
$$p(s) = \frac{2}{\pi}\frac{n_{\mathrm{eff}}}{\sigma_{\mathrm{eff}}}\, s\, e^{-\frac{s^2}{\pi}} \left(1 - e^{-\frac{s^2}{\pi}}\right)^{n_{\mathrm{eff}}-1}.$$

The probability of false detection as a function of SNR in VGaPS experiments

SNR	$P_f(s)$
4.96	0.3
5.19	0.1
5.61	0.01
5.99	0.001
6.68	10^{-5}

Computation of totals

- Compute Scan Reference Time



- Compute geocentric a priori path delay and delay rate on SRT.

- Extrapolate the a priori path delay if needed. Compute VTD path delay and extrapolate the difference: used a priori minus VTD.

- Compute clock function and atmospheric path delay **used** in the a priori model.

- Compute epochs the the reference station moment of time using iterations:

$$t_{\mathrm{srt,ref1}} = t_{\mathrm{srt,gc}} - \tau(t_{\mathrm{srt,ref1}}) + \tau_{\mathrm{cl1}} + \tau_{\mathrm{at1}}$$

$$t_{\mathrm{srt,ref2}} = t_{\mathrm{srt,gc}} - \tau(t_{\mathrm{srt,ref2}}) + \tau_{\mathrm{cl2}} + \tau_{\mathrm{at2}}$$

- Compute the a priori reference station based delay, phase, and phase delay rate:

$$\tau_{\mathrm{apr,ref}} = \tau_{\mathrm{apr,gc}}(t_{\mathrm{srt,ref2}}) - \tau_{\mathrm{apr,gc}}(t_{\mathrm{srt,ref1}}) + (\tau_{\mathrm{cl2}} - \tau_{\mathrm{cl1}})$$

- Re-adjust a priori phase for change of the frequency.

- Compute total delay:

$$\tau_{\mathrm{gr}} = \tau_{\mathrm{apr,ref}} + s \cdot \tau_{\mathrm{res}} + s \cdot \dot{\tau}_{gr}\,(t_{\mathrm{srt}} - t_{\mathrm{frt}})$$

where s is the baseline sign, $+1$ or -1.