Post-correlation Analysis & Fringe-fitting

2nd IVS VLBI School – Hartebeesthoek, SA

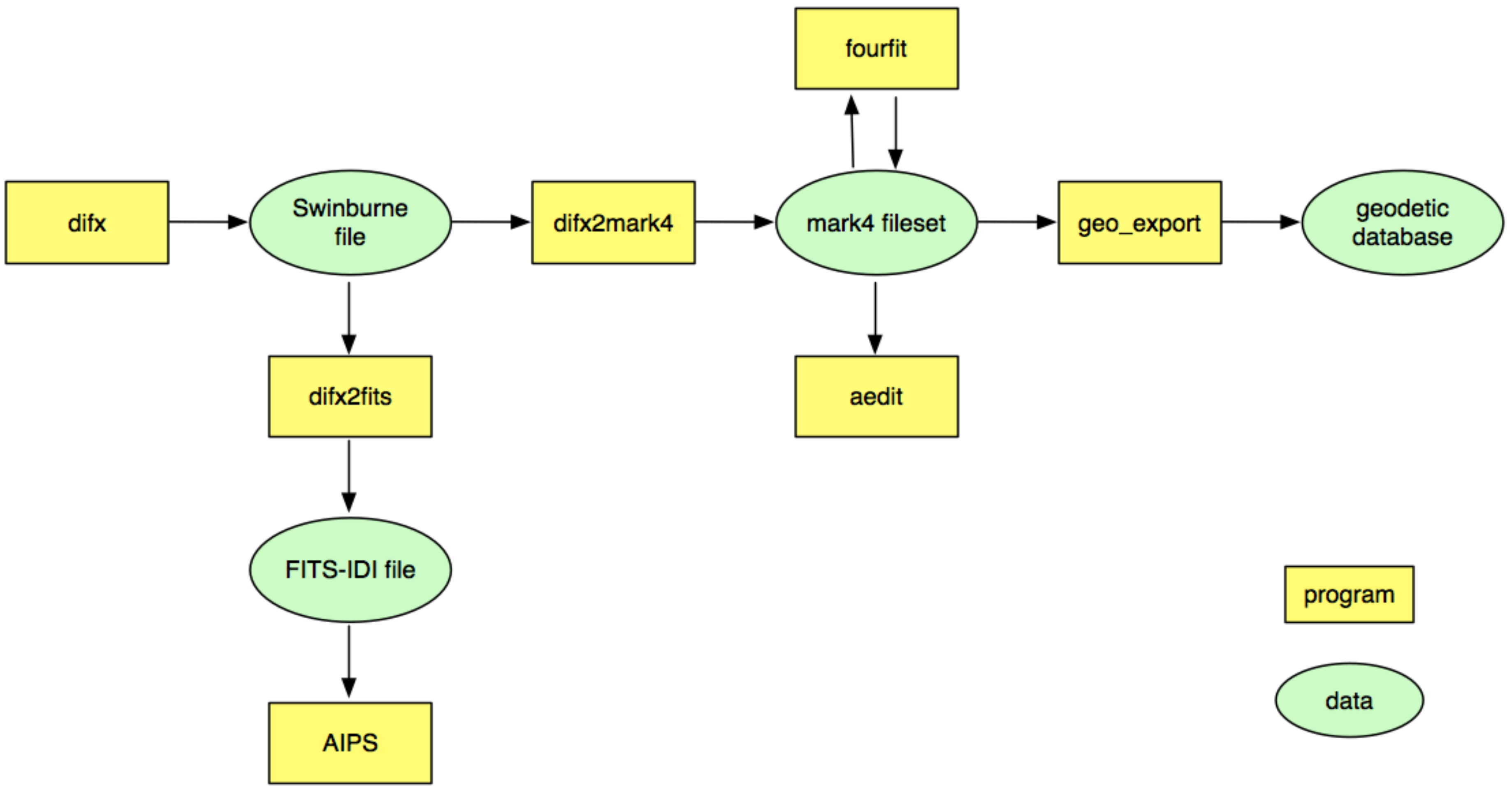
Roger Cappallo
MIT Haystack Observatory

2016.3.11

overview

- fringe-fitting
    - theory
    - practical example within fourfit
- data quality analysis
    - key to successful operations
- data export to geodetic databases

typical processing dataflow



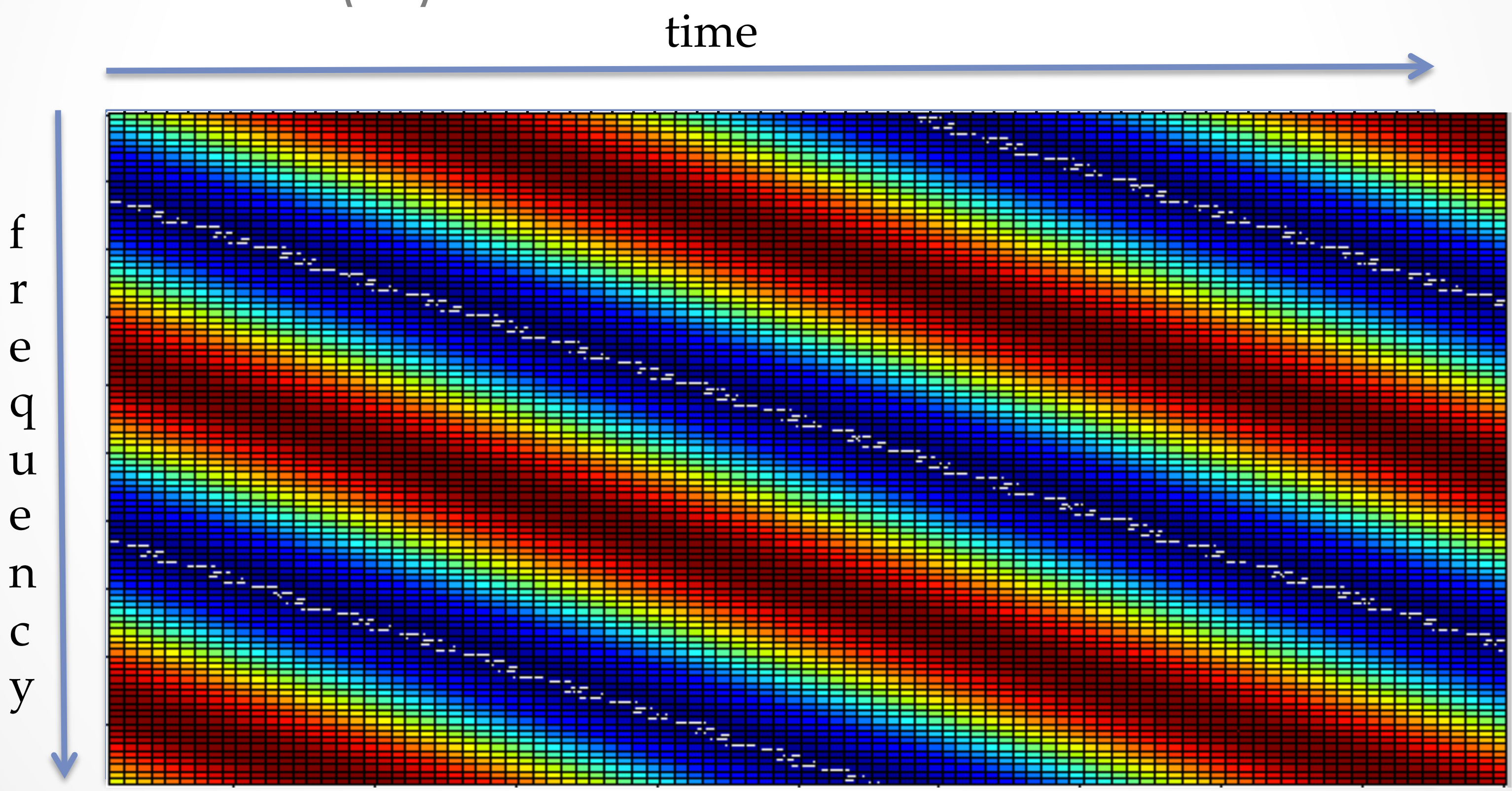
why is fringe-fitting even necessary?

- Correlator model is good, but not perfect
- Typically, antenna models and locations are now very good, but...
- Source positions are imperfect, and can vary with time and frequency
- Atmosphere and ionosphere are time-variable and unpredictable
- GPS clock information has significant errors at the VLBI level of accuracy

Fringe-fitting removes remaining non-random signatures by incremental changes to the correlation parameters

central concept of fringe-fitting

- correlator produces a 2-D complex array of visibilities $\mathcal{V}$ (f,t)



typical patterns in visibilities

- mean amplitude
- quasi-linear drift of phase with time
- quasi-linear drift of phase with frequency

(all trends have noise added to them, often dominant)

extracted parameters

- principally for astronomy:
  - ρ amplitude
  - Φ phase

  } visibility➔FT➔image

- principally for geodesy:
  - τ_g group delay: variation of phase with frequency
  - $\dot{\tau}_g$ delay-rate: rate of change of τ_g, derived from the variation of phase with time

- nuisance (at least for us)
  - ΔTEC: differential Total Electron Content (of ionosphere)

snr example

- signal-to-noise ratio for a scan is given by:

$$snr_{scan} = snr_{sample} \times \sqrt{(\#samples)}$$

- where:
  - $snr_{sample} = \rho$
  - #samples = 2 B T

- VGOS example 4 GS/s * 30s = 1.2e11 samples, so snr increases by a factor of ~ 10^5
- this turns a 0.001 correlation into a scan snr of 300
- for typical scans a minimum snr of 6 or 7 is required for detection

Fourier Transform

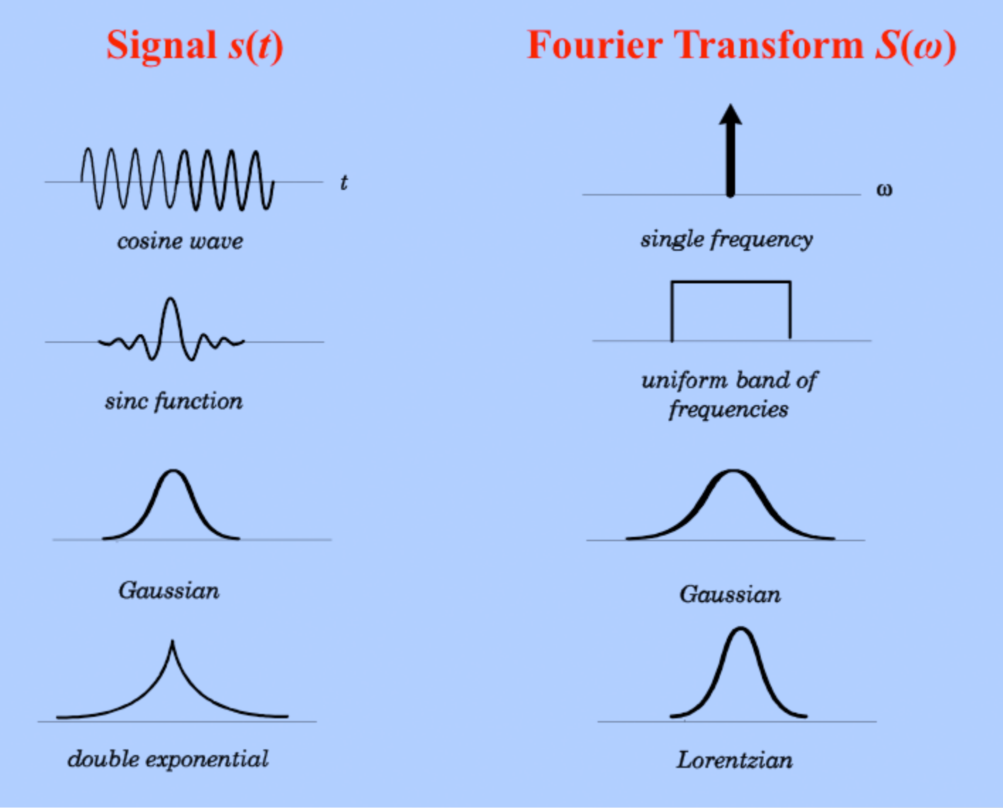
- generates alternative representation of a function in its conjugate domain (e.g. time ~ frequency)

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega$$

- dft
- fft

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i 2\pi k \frac{n}{N}} \qquad k = 0, \dots, N-1$$



fringe-fitting in *fourfit*

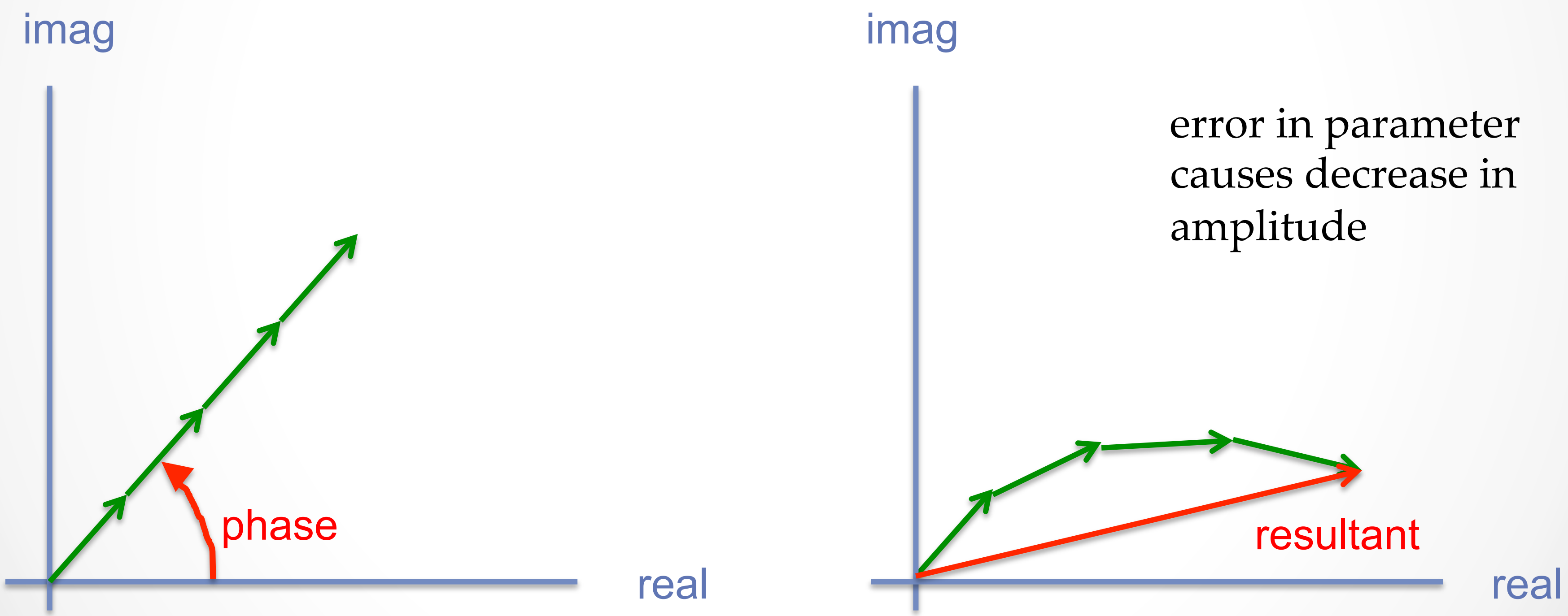
- 2 steps
  - coarse grid search
  - refinement of parameter estimates
- grid search done via FFT's:
  - over frequency to find delay
  - over time to find fringe/delay rate
  - over "lag" to find single-band delay
- refinement
  - counter-rotate data and coherently sum:

$$\mathbf{g}(\tau, \dot{\tau}) = \sum_f \sum_t \boldsymbol{V}(f, t) e^{-2\pi i(f\dot{\tau}t + f\tau + \delta\phi)}$$

  - interpolate from closely-spaced grid-points

coherent addition of visibilities

(idealized noiseless case)



error in parameter causes decrease in amplitude

$$\mathbf{g}(\tau, \dot{\tau}) = \sum_f \sum_t \boldsymbol{V}(f, t) e^{-2\pi i (f \dot{\tau} t + f \tau + \delta\phi)}$$

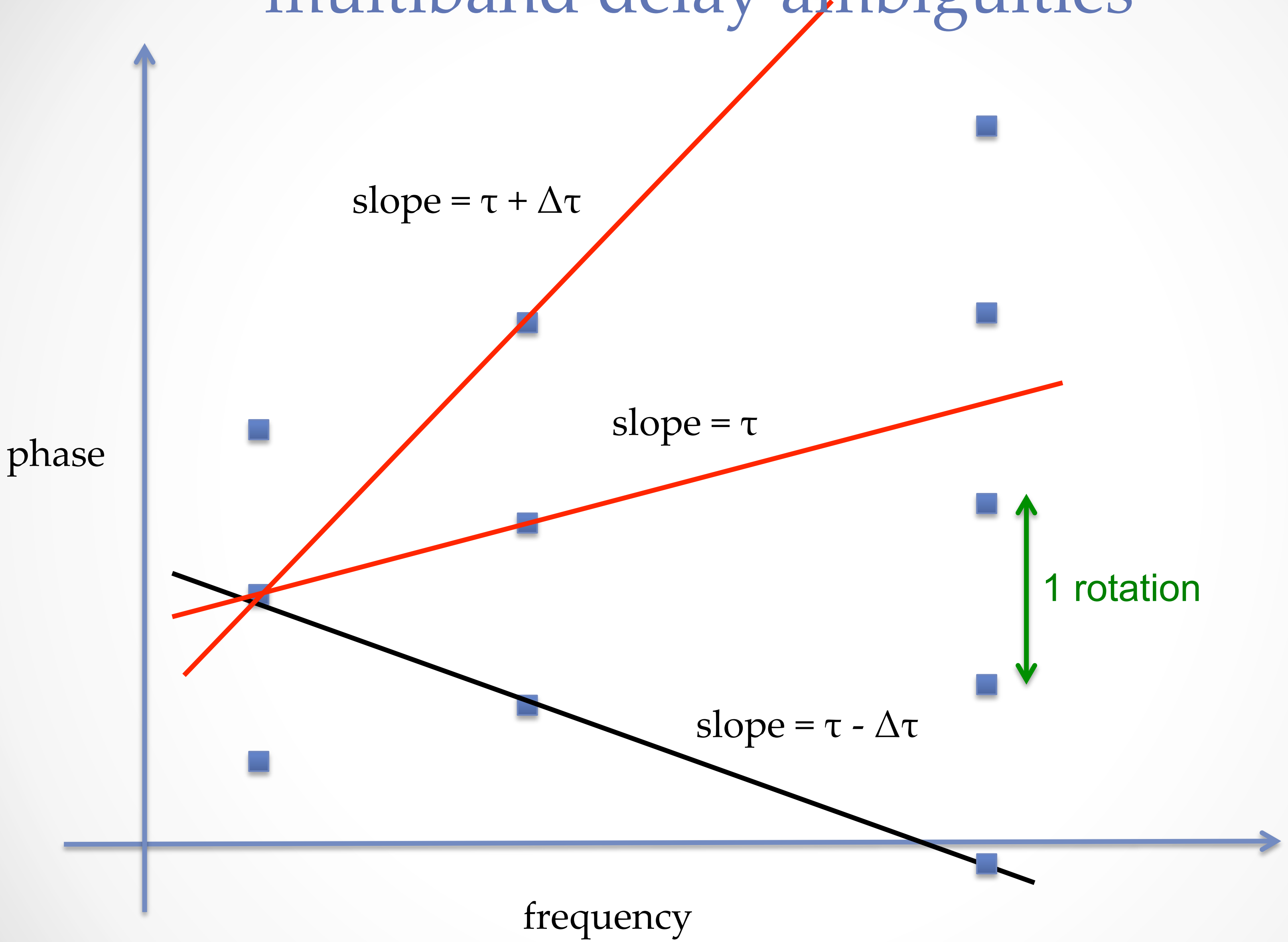
singleband delay

- slope of phase vs. frequency within one channel (e.g. 32 MHz, for VGOS)
- determined on a baseline by estimating slope of phase vs. freq. of a radio source
- instrumental contribution for a single antenna can be found by using the phases of phase-cal tones embedded in the channel

multiband delay

- determined by "collapsing" each channel's data down to a single phase per channel, and then finding the slope of those phases across their frequency range
- by spacing channels apart, a wider range of frequencies is covered, leading to a more accurate slope
- technique is called "bandwidth synthesis"
- ambiguities are unavoidable...

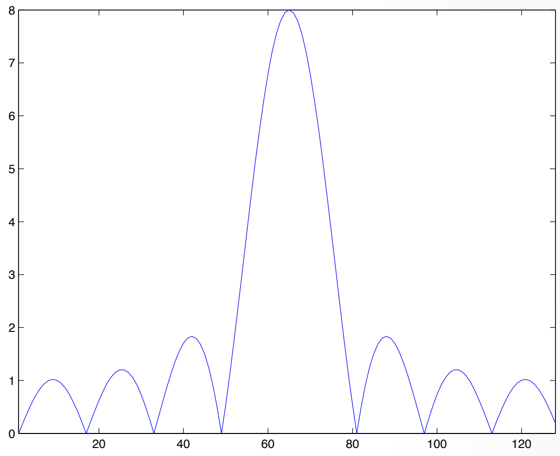
multiband delay ambiguities



delay resolution function

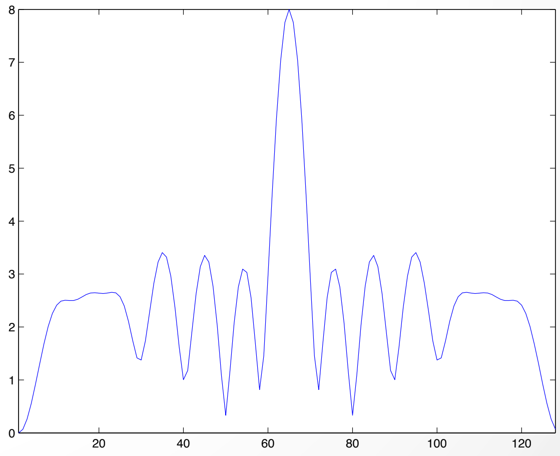
FFT of cross-power spectrum

freqs = [1 2 3 4 5 6 7 8]



freqs = [1 2 5 8 10 13 14 15]

has narrower peak, higher sidelobes



(cf. Arsac arrays & Golomb rulers)

multiband delay vs. singleband delay

- different due to things that affect single channels or groups of channels (e.g. cable lengths, filter delays)
- by correcting channels with pcal-derived delays there is hope to go to (just) multiband delay
- ambiguity spacing in delay is inverse of greatest common frequency difference
  - VGOS mbd: 32 MHz spacing ➔ 31.25 ns ambiguity
  - VGOS sbd: 128 spectral pts/channel ➔ 1/8 MHz spacing ➔ 8 μs ambiguity

delay-rate vs. fringe-rate

$\frac{d\tau}{dt}$ **delay-rate** (group-delay rate) is rate of change of delay, and is dimensionless

$\frac{d\phi}{dt}$ **fringe-rate** is rate of change of fringe phase, typically in Hz or mHz. It is the differential Doppler-shift

related by $$\frac{d\tau}{dt} = \frac{1}{f} \times \frac{d\phi}{dt}$$

overcoming instrumental shortcomings

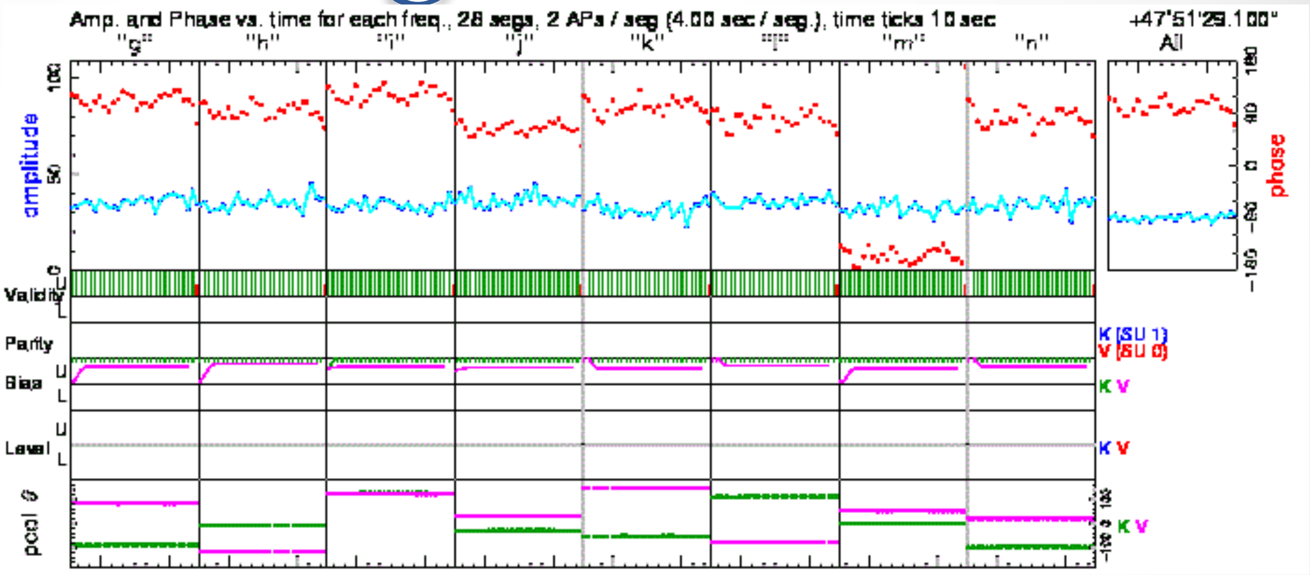
- GPS clock errors
  - search over range of delays
- RFI
  - delete channels or times as necessary
- data defects (e.g. off-source)
  - delay start or stop of fit
- phase & delay (mis)calibration
  - use pcal tones and/or manually adjust delays and pcal phases

phase calibration

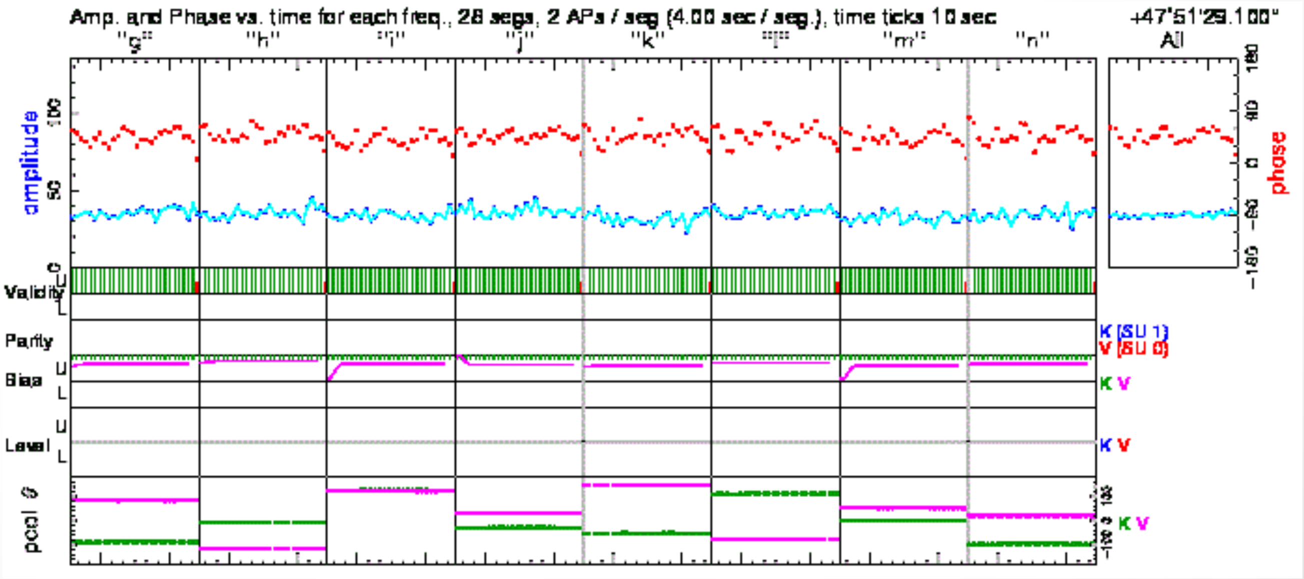
- legacy "normal"
  - 1 tone per channel
  - deprecated!
- multitone
  - many tones per channel
  - capability to correct channel delay
- manual
  - typically set to constant values for whole experiment
  - line up phases with strong calibrator source
  - slight tweaks just change the clock estimate

phase-cal aligns channels

before:

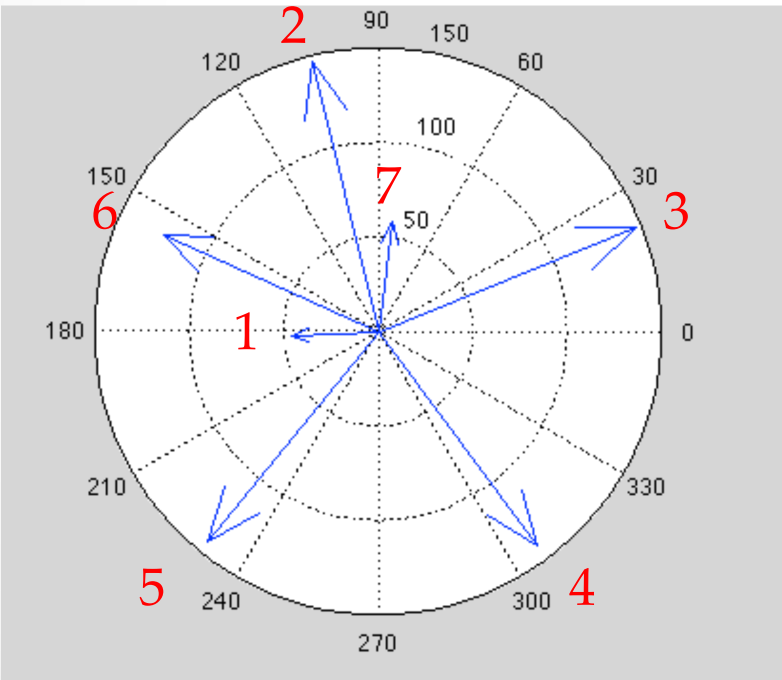


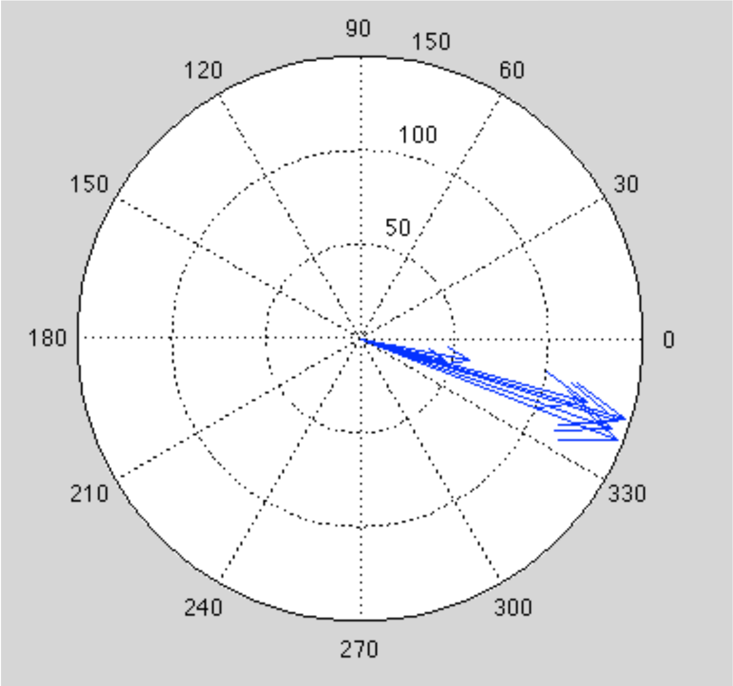
after:



channel-based delay correction using multitone pcal

$$V^*(f_k) = V(f_k)e^{-2\pi i(f_k \delta\tau)}$$

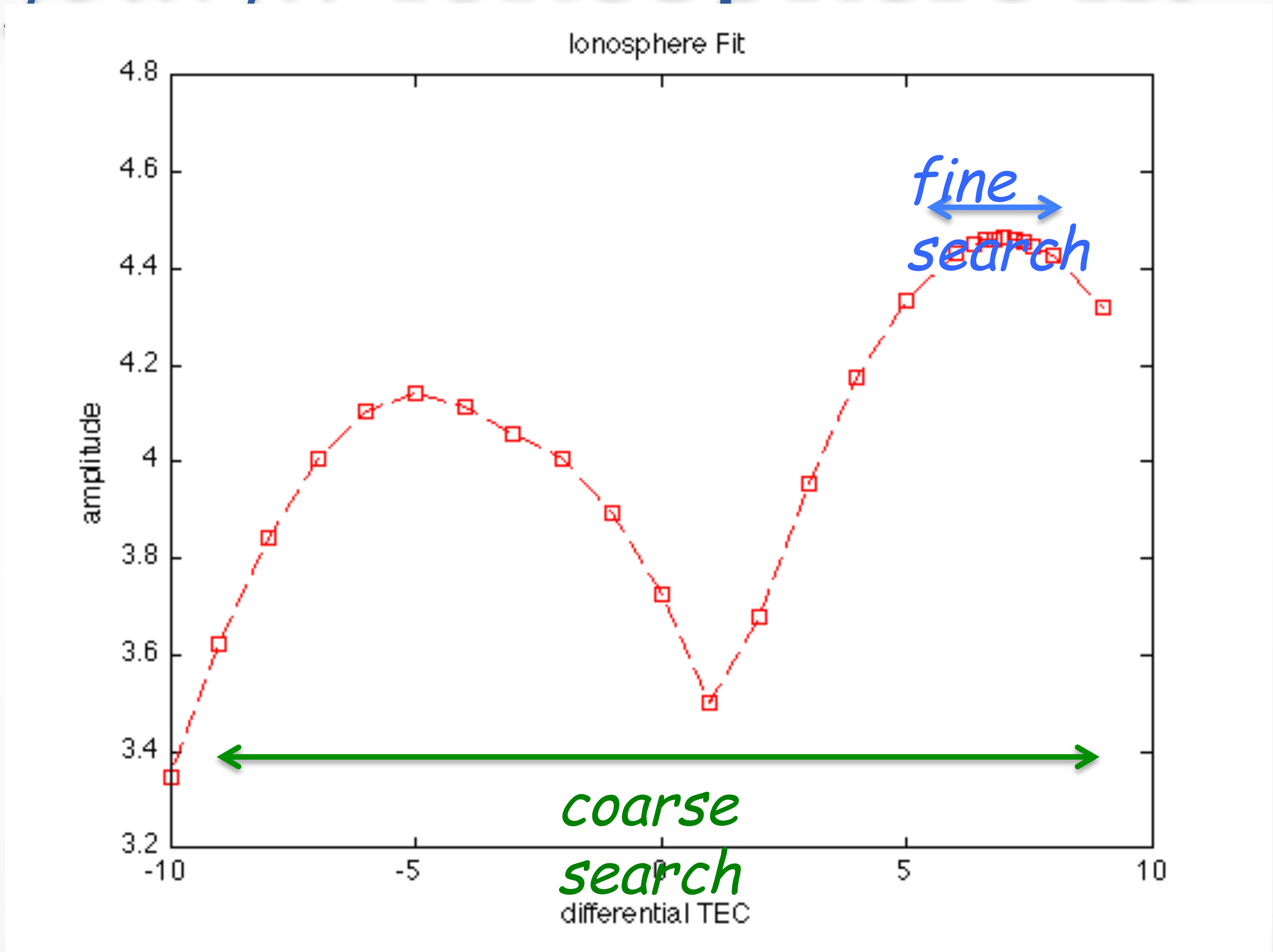




ionosphere

- phase of each freq channel affected by differential path integral of charges (Total Electron Content)
- 1 TEC unit = 10^{16} electrons / m^2
- $\Delta \phi = c \times \Delta TEC / f$
- differential TEC can be fit and/or specified *a priori*
  - all-sky models from GPS available, but not yet used
  - fit made difficult by nonlinearity
  - search for peak of coherent sum of all bands
- ionosphere and group delay estimate are strongly correlated ~92% for VGOS

fourfit ionosphere fit



combining linear polarizations in fourfit

- Maximize sensitivity in τ_g by combining all 4 Stokes polarization products
- Form an approximation to Stokes I:
  - from the 4 correlation products form

$$I \cong (HxH + VxV) \cos\Delta + (HxV - VxH) \sin\Delta$$

Δ = differential parallactic angle

  - correct to first order in the D terms
- Also have mixed combinations to legacy stations

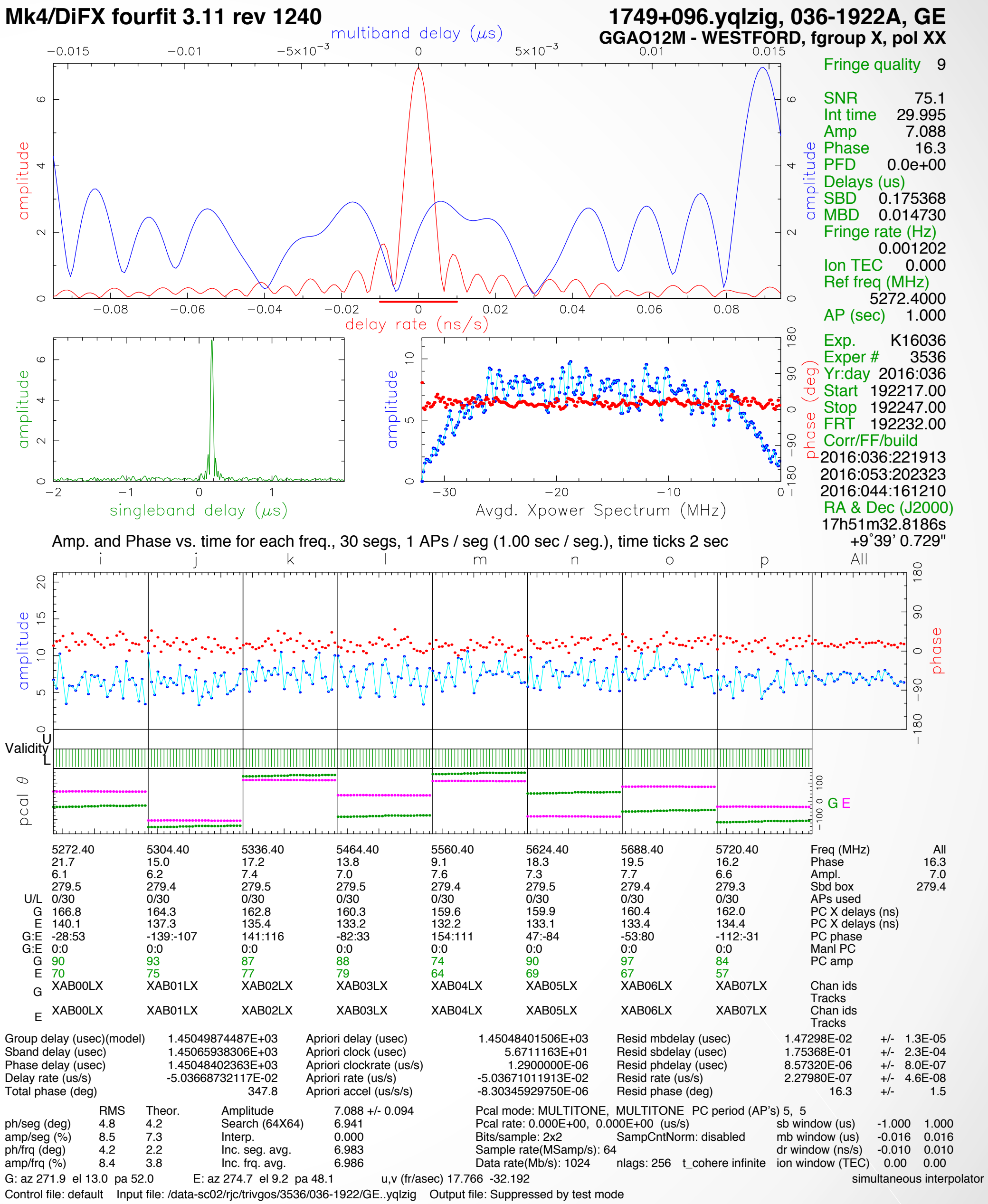
e.g. {RxV, RxH, LxV, LxH}

fourfit output

- "mk4" data file
  - used for fourfit input as well
  - set of files tied by a common suffix
    - type 0: root file, contains vex statements for scan
    - type 1: correlator output (visibilities), 1 per baseline
    - type 2: fourfit output, per baseline & by ff fit
    - type 3: station files, 1 per station
- fringe plot
  - single page w/ graphical and printed summary

fringe-plot example

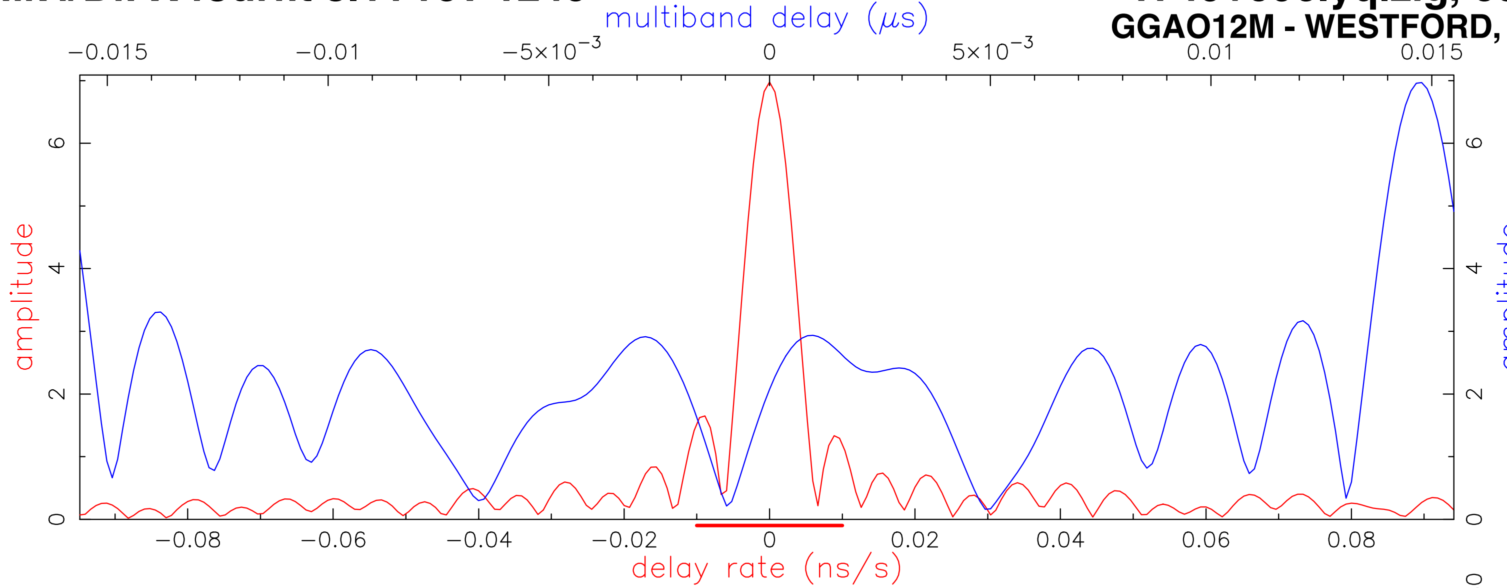
- concise summary, but crowded
- plots
  - multiband delay
  - singleband delay
  - delay-rate
  - cross-power spectrum
  - phase & amp by channel as function of time
  - pcal amp & phase(t)
  - data fractions
- text
  - residual fit parameters
  - total values
  - metadata
  - statistics



Mk4/DiFX fourfit 3.11 rev 1240

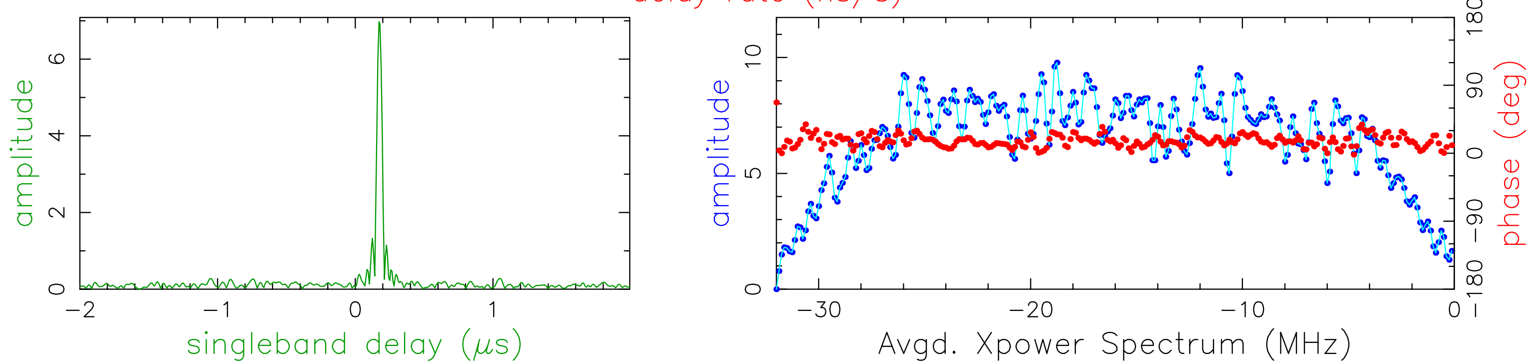
1749+096.yqlzig, 036-1922A, GE

GGAO12M - WESTFORD, fgroup X, pol XX



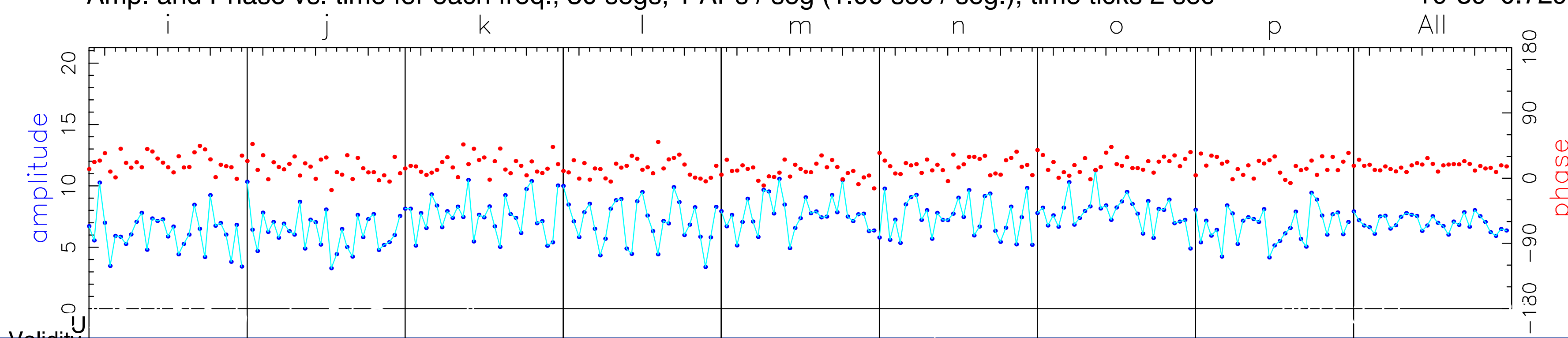
Fringe quality 9

SNR 75.1
Int time 29.995
Amp 7.088
Phase 16.3
PFD 0.0e+00
Delays (us)
SBD 0.175368
MBD 0.014730
Fringe rate (Hz)
0.001202
Ion TEC 0.000
Ref freq (MHz)
5272.4000
AP (sec) 1.000



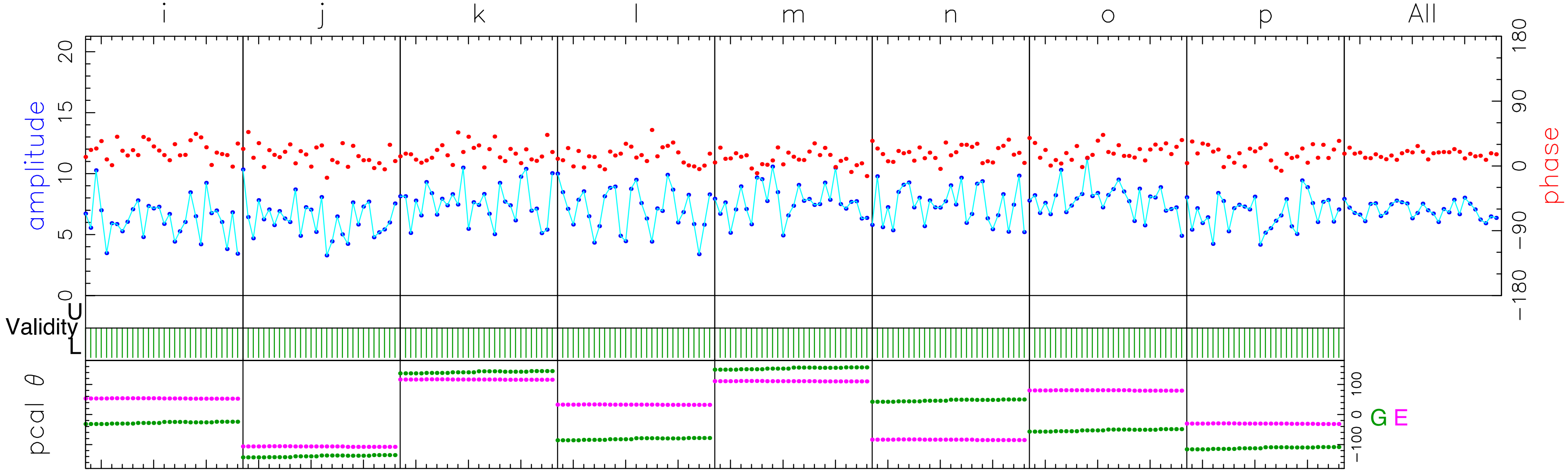
Exp. K16036
Exper # 3536
Yr:day 2016:036
Start 192217.00
Stop 192247.00
FRT 192232.00
Corr/FF/build
2016:036:221913
2016:053:202323
2016:044:161210
RA & Dec (J2000)
17h51m32.8186s
+9°39' 0.729"

Amp. and Phase vs. time for each freq., 30 segs, 1 APs / seg (1.00 sec / seg.), time ticks 2 sec



17h51m32.8186s
+9˚39' 0.729"

Amp. and Phase vs. time for each freq., 30 segs, 1 APs / seg (1.00 sec / seg.), time ticks 2 sec



	i	j	k	l	m	n	o	p		All
	5272.40	5304.40	5336.40	5464.40	5560.40	5624.40	5688.40	5720.40	Freq (MHz)	All
	21.7	15.0	17.2	13.8	9.1	18.3	19.5	16.2	Phase	16.3
	6.1	6.2	7.4	7.0	7.6	7.3	7.7	6.6	Ampl.	7.0
	279.5	279.4	279.5	279.5	279.4	279.5	279.4	279.3	Sbd box	279.4
U/L	0/30	0/30	0/30	0/30	0/30	0/30	0/30	0/30	APs used	
G	166.8	164.3	162.8	160.3	159.6	159.9	160.4	162.0	PC X delays (ns)	
E	140.1	137.3	135.4	133.2	132.2	133.1	133.4	134.4	PC X delays (ns)	
G:E	-28:53	-139:-107	141:116	-82:33	154:111	47:-84	-53:80	-112:-31	PC phase	
G:E	0:0	0:0	0:0	0:0	0:0	0:0	0:0	0:0	Manl PC	
G	90	93	87	88	74	90	97	84	PC amp	
E	70	75	77	79	64	69	67	57		
G	XAB00LX	XAB01LX	XAB02LX	XAB03LX	XAB04LX	XAB05LX	XAB06LX	XAB07LX	Chan ids Tracks	
E	XAB00LX	XAB01LX	XAB02LX	XAB03LX	XAB04LX	XAB05LX	XAB06LX	XAB07LX	Chan ids Tracks	

Group delay (usec)(model)	1.45049874487E+03	Apriori delay (usec)	1.45048401506E+03	Resid mbdelay (usec)	1.47298E-02	+/-	1.3E-05
Sband delay (usec)	1.45065938306E+03	Apriori clock (usec)	5.6711163E+01	Resid sbdelay (usec)	1.75368E-01	+/-	2.3E-04
Phase delay (usec)	1.45048402363E+03	Apriori clockrate (us/s)	1.2900000E-06	Resid phdelay (usec)	8.57320E-06	+/-	8.0E-07
Delay rate (us/s)	-5.03668732117E-02	Apriori rate (us/s)	-5.03671011913E-02	Resid rate (us/s)	2.27980E-07	+/-	4.6E-08
Total phase (deg)	347.8	Apriori accel (us/s/s)	-8.30345929750E-06	Resid phase (deg)	16.3	+/-	1.5

	RMS	Theor.		
ph/seg (deg)	4.8	4.2	Amplitude	7.088 +/- 0.094
amp/seg (%)	8.5	7.3	Search (64X64)	6.941
ph/frq (deg)	4.2	2.2	Interp.	0.000
amp/frq (%)	8.4	3.8	Inc. seg. avg.	6.983
			Inc. frq. avg.	6.986

Pcal mode: MULTITONE, MULTITONE PC period (AP's) 5, 5
Pcal rate: 0.000E+00, 0.000E+00 (us/s)
Bits/sample: 2x2 SampCntNorm: disabled
Sample rate(MSamp/s): 64
Data rate(Mb/s): 1024 nlags: 256 t_cohere infinite

sb window (us)	-1.000	1.000
mb window (us)	-0.016	0.016
dr window (ns/s)	-0.010	0.010
ion window (TEC)	0.00	0.00

G: az 271.9 el 13.0 pa 52.0 E: az 274.7 el 9.2 pa 48.1 u,v (fr/asec) 17.766 -32.192 simultaneous interpolator

Control file: default Input file: /data-sc02/rjc/trivgos/3536/036-1922/GE..yqlzig Output file: Suppressed by test mode

fourfit control files

- text files with simple syntax
- there are ~95 keywords known to *fourfit*
- syntactic elements
  - if, and, or, not, <, >, ?
- data selectors
  - station, baseline, source, scan, f_group
- filtering
  - freqs, start, stop, etc.
- corrections
  - pc_mode, pc_phases, ionosphere, ref_freq, lsb_offset, etc.
- search control
  - sb_win, mb_win, dr_win, ion_win, etc.

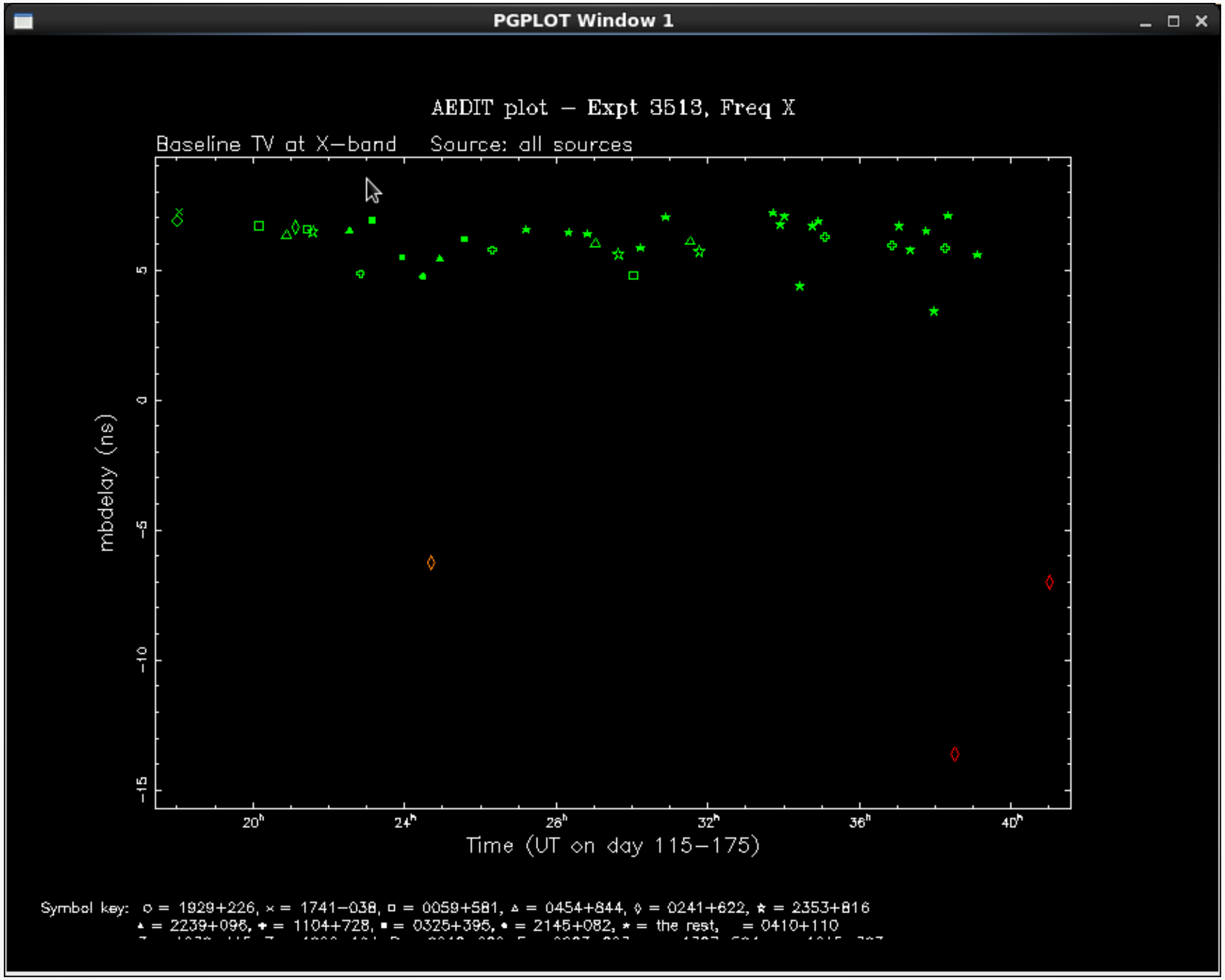
example control file

```
ref_freq  8213.15                    * global commands come first

start -10
if scan 288-210210
   sb_win .37 .37

if scan > 289-132510                 * don't use any scans after 1325
   skip true

if station L and f_group X
   freqs a+ b c d- e f g h

if station L and f_group S
  pc_mode manual
  pc_phases ijkmn 4.5 -78 39 +12 0
if station A
   pc_mode multitone
   pc_period 30
   pc_tonemask abcdefgh 0 0 8 0 4 0 5 0
   pc_phases_l abcdefgh 12 13 11 12 24 -6 38 110
   pc_phases_r abcdefgh 11 29 14 11 64 -2 44 132
   samplers 2 abcd efgh
   pc_delay_l 30.2  pc_delay_r -5.9
   ionosphere 18.0
if station V or baseline KT and source 3C279
   sb_win -0.5 0.5    mb_win 0.02 0.02  dr_win -1.0E-6 0.5E-6
```

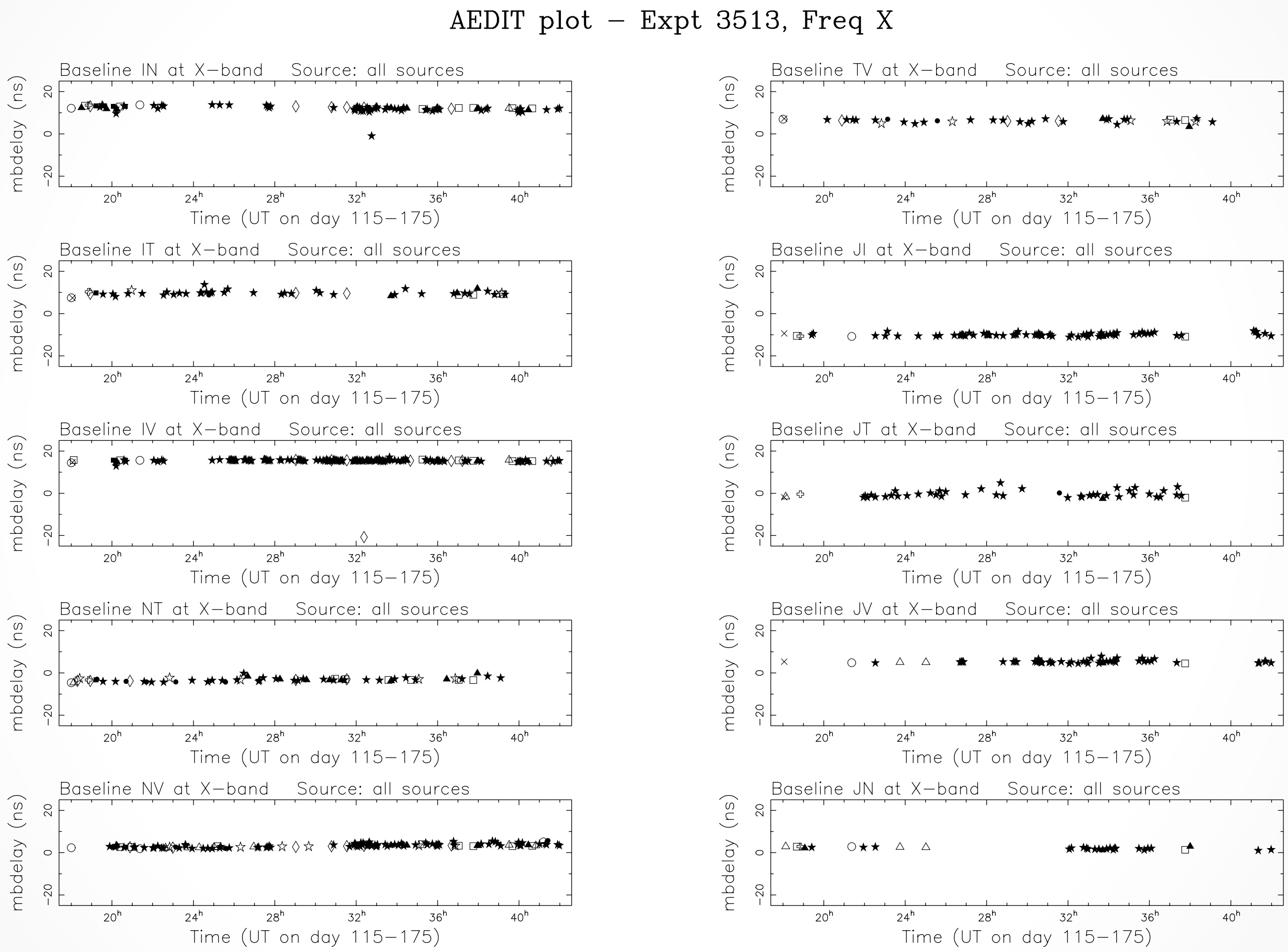
aedit

- name comes from “A” files, which are text files with one (very long) line summaries of fringe-fits
- used for
  - data quality analysis of the correlation & post-processing output
  - selection and filtering of data to be sent on to the geodetic databases

mbd residual screen-shot



mbd residual print-out



Symbol key: ○ = 1929+226, × = 1741–038, □ = 1348+308, △ = 1736+324, ◊ = 0454+844, ☆ = 1104+728
▲ = 1022+194, ✚ = 1520+319, ■ = 1039+811, • = 0325+395, ★ = the rest

fourfit quality codes

Aedit: Quality Code

igg | universitätbonn

QC = 0 Fringes not detected (PFD > 1e-4).

= 1-9 Fringes detected, no error condition. Higher number => better quality.

= B Interpolation error in fourfit.

= D No data in one or more freq. channels.

= E Max fringe amplitude at the edge of SBD, MBD or DR window.

= F "Fork" problem in processing.

= G Fringe amp. in one or more channels is < 0.5 mean amp. (for SNR > 20).

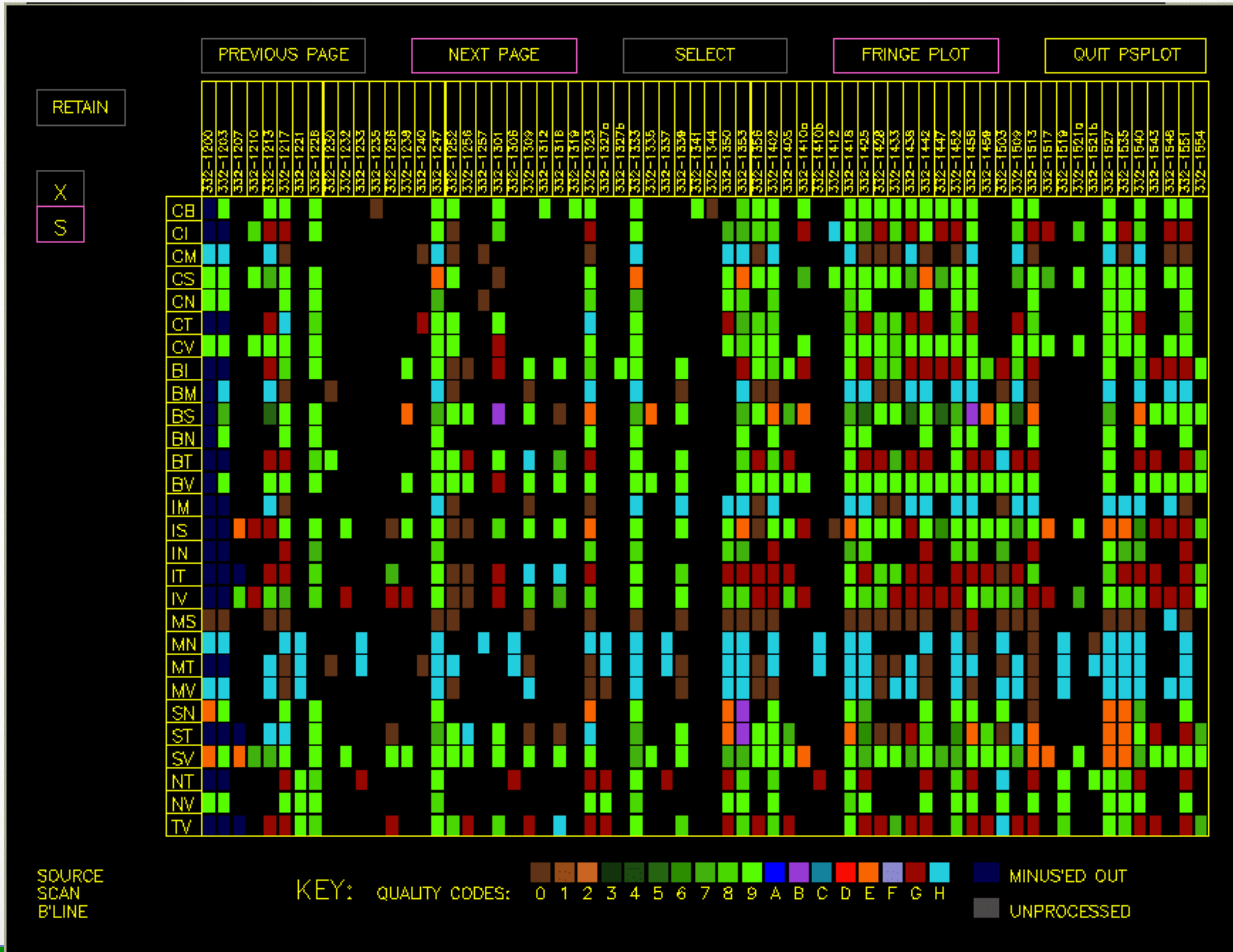
= H Low pcal-amplitude.

= N No valid correlator data.

56

A. Bertrarini

aedit's psplot feature



database export

- Several methods used
- Haystack
  - Legacy S/X: tarball of fourfit output fileset (roots & type 2's) sent to Goddard, where they are imported
  - VGOS: preliminary program called vgosDBMake
- Bonn
  - geo_export run, invoking calc/solve to create database

summary

- fringe-fitting
  - why it is necessary
  - fourfit – its theory, implementation, and interpretation
- data quality analysis
  - aedit
- data export to geodetic databases

thank you!