

DiFX Correlation & Post-Correlation Analysis

Alessandra Bertarini

IGG University of Bonn & MPIfR Bonn

- DiFX -> Distributed FX correlator.
- DiFX is a software correlator.
- DiFX is a free downloadable software from:
<http://cira.ivec.org/dokuwiki/doku.php/difx/installation>
- DiFX needs IPP libraries (IPP requires licence).



RAIDs

RAIDs +
fxmanager

nodes +
frontend and
frontend2

nodes

60 compute nodes

8 cores per node

20 Gbps
InfinBand

Network cards

10 RAIDs
(220 TB)

1 service node
(with keyboard &
monitor)

2 user interaction
nodes (frontend
& frontend2)

DiFX is software running on various computer clusters.
Every cluster performance is different, but...

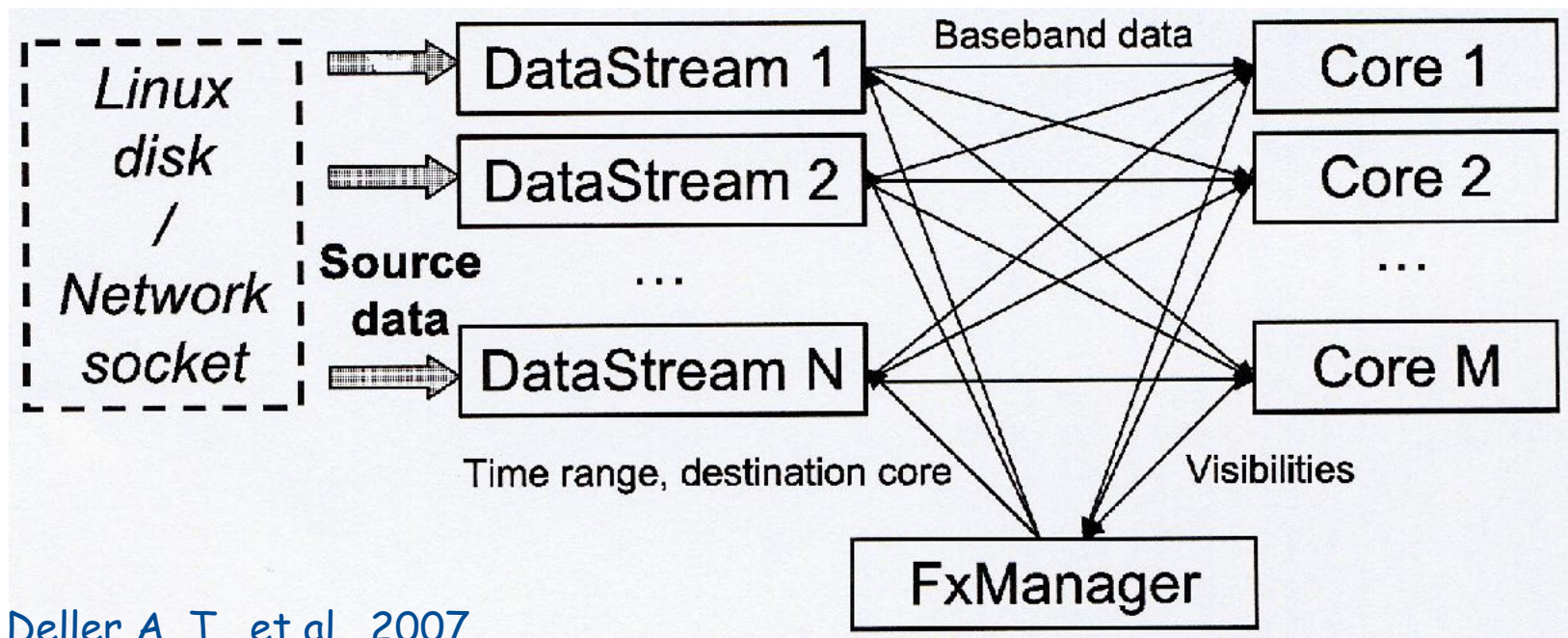
the fundamental operations performed by the correlator
are the same.

DiFX: receives digitized signals
 applies the correlator model
 pads the data from 2 bits to 16 bits
 aligns the data within +/- 1 sample
 performs an FFT
 performs a fractional-sample delay correction
 performs a complex multiplication & integrates
 writes the complex visibilities (in freq. domain)

Mark 5 connected to DiFX via Ethernet or InfiniBand

FxManager controls operation: send data from data manager nodes (DataStream) to processing nodes (Cores).

Processed data from Cores sent back to FxManager.



Deller A. T., et al., 2007

DiFX reads data in: Mark 4/VLBA format,
Mark 5B format,
VDIF format (single thread),
LBA format

DiFX needs one valid VEX file and a v2d file (vex-to-DiFX).

Data quality control: AIPS (too complex for the purpose)
HOPS fourfit

VEX = (VLBI EXperiment) file

VEX files are used from the stations for:

- Sky Frequency
- LO tuning
- Recording speed
- Polarization
- No. of BBCs
- Sources to be observed
- Length of the scans
- Track assignment

VEX files are used by the correlators for:

- Sky Frequency → relevant for fringe rotator
- LO tuning → relevant for fringe rotator
- Recording speed → relevant for playback speed
- Polarization → relevant for channel assignment
- No. of BBCs → relevant for channel assignment
- Sources to be observed → coordinates for corr. model
- Length of the scans → relevant for playback
- Track assignment → relevant for channel assignment
- Antenna coordinates (not required for observing)

Correlator's VEX files need extra information:

Earth rotation parameters (x-wobble, y-wobble and UT1)

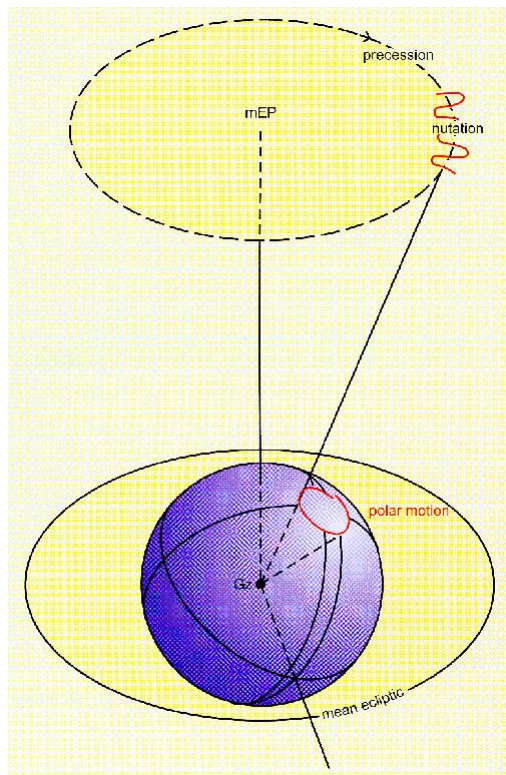
Clock information (gps-fmout from field system logs)

Data source (Mark 5 module, files on RAID)

Correlator's VEX files need (sometimes) to be changed:

Track assignment (only tape-like tracks are present in VEX)

Earth Orientation Parameters (x-wobble, y-wobble and UT1)



The predicted values are published from
US Naval Observatory:

http://128.183.20.176/solve_save/usno_finals.erp

DiFX ancilliary program `geteop.pl` read
the USNO file, reformat it and creates a
file called `EOP.txt`

EOP: VEX example for observation on DOY 035.

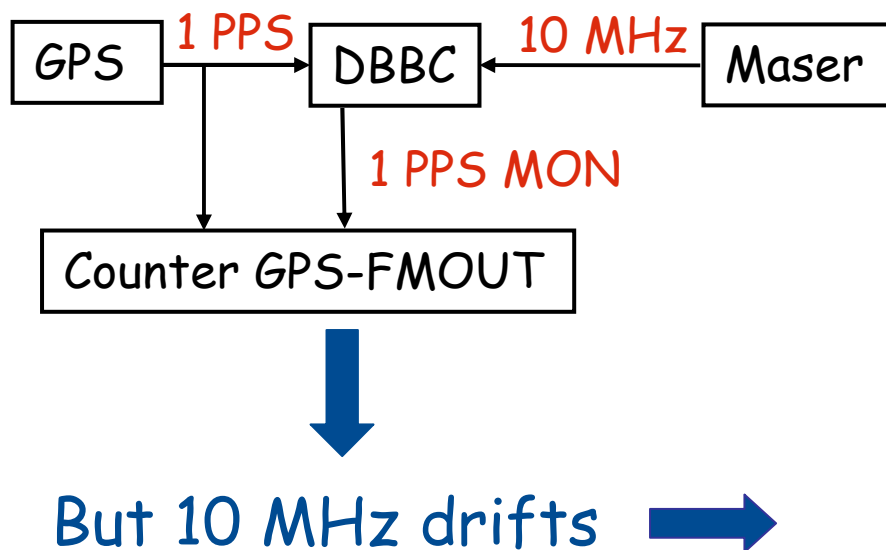


11

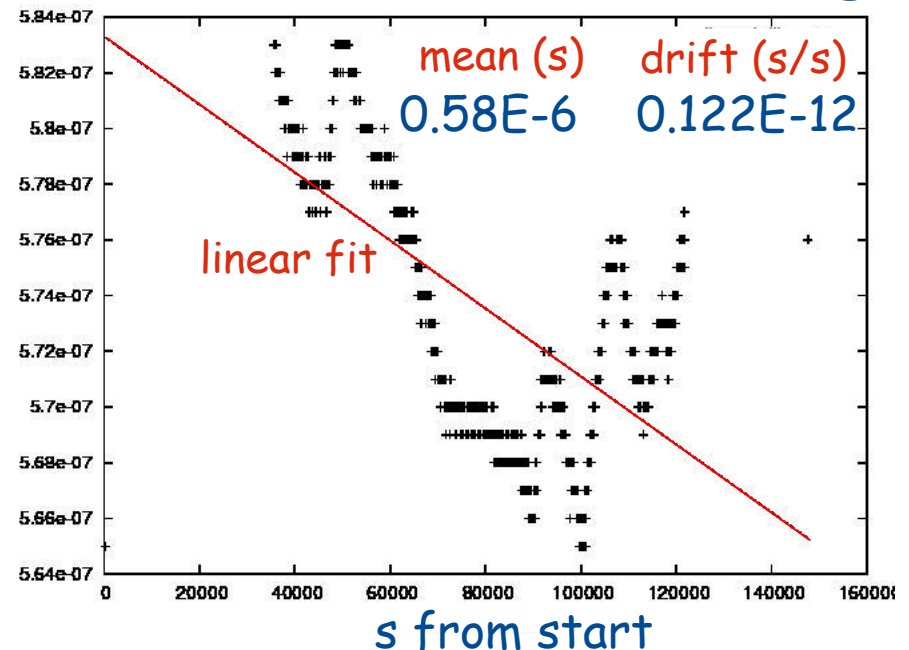
```
$EOP;  
def EOP0;  
  TAI-UTC= 35 sec;  
  A1-TAI= 0 sec;  
  eop_ref_epoch=2013y033d;  
  num_eop_points=1;  
  eop_interval=24 hr;  
  ut1-utc   = 0.237134 sec;  
  x_wobble  = 0.042530 asec;  
  y_wobble  = 0.313450 asec;  
enddef;  
def EOP4;  
  [...]  
enddef;
```

Note: DiFX needs EOPs for 5 days of which two prior to the observation !

CLOCK: estimates the time difference between the data time stamps (from formatter/M5B/ FiLa 10G) and UTC coming from GPS.



GPS-FMOUT from FS log



```

$CLOCK;
def Ny;
  clock_early=2012y142d17h00m : X usec :2012y142d17h00m0s : x ;
enddef;
def Wz;
  clock_early=2012y142d17h00m : Y usec :2012y142d17h00m0s : y;
enddef;
[...]
```

13

The **X** and **Y** μ s are the mean gps-fmout "clock" values.

The **x** and **y** s/s are the clock drifts.

DiFX needs to know whether the data are on a Mark 5 module and need to know which module.

```
$TAPELOG_OBS;  
  def Kk ;  
    VSN=1 : HOB+0025 : 2013y043d17h00m00s :2013y044d16h57m46s ;  
  enddef ;  
  def Tc ;  
    VSN=1 : HART-014 : 2013y043d17h00m00s :2013y044d16h17m56s ;  
  enddef ;
```

Note: E-transfered stations do not appear in the VEX !

Check "track" assignment: VEX speaks (still) tape language!

Mk 4	VSI=geo	VSI=astro	Mk 4	VSI=geo	VSI=astro
1US	0	0	1LS	16	16
1UM	1	1	1LM	17	17
2US	2	2	2LS	-	18
2UM	3	3	2LM	-	19
3US	4	4	3LS	-	20
3UM	5	5	3LM	-	21
4US	6	6	4LS	-	22
4UM	7	7	4LM	-	23
5US	8	8	5LS	-	24
5UM	9	9	5LM	-	25
6US	10	10	6LS	-	26
6UM	11	11	6LM	-	27
7US	12	12	7LS	-	28
7UM	13	13	7LM	-	29
8US	14	14	8LS	18	30
8UM	15	15	8LM	19	31

Check "track" assignment: VEX speaks (still) tape language!

Mk 4	VSI=geo	VSI=astro
9US	21	-
9UM	22	-
10US	23	-
10UM	24	-
11US	25	-
12UM	26	-
12US	27	-
13UM	28	-
13US	29	-
14UM	30	-
14US	31	-

In VEX enter VSI output + 2 !

i.e. 1US: VSI output = 0 → VEX TRACK = 0 + 2 = 2

TRACKS sorting in VEX:

```

$FREQ;
def GEOSX-SX01;
chan_def = &X:8212.99 MHz:U:8.000 MHz:&CH01:&BBC01:&U_cal;
[.]
enddef;
[.]
$TRACK;
def Mark5B;
fanout_def = A : &CH01 : sign : 1 : 02 ;
[.]
enddef;
    
```

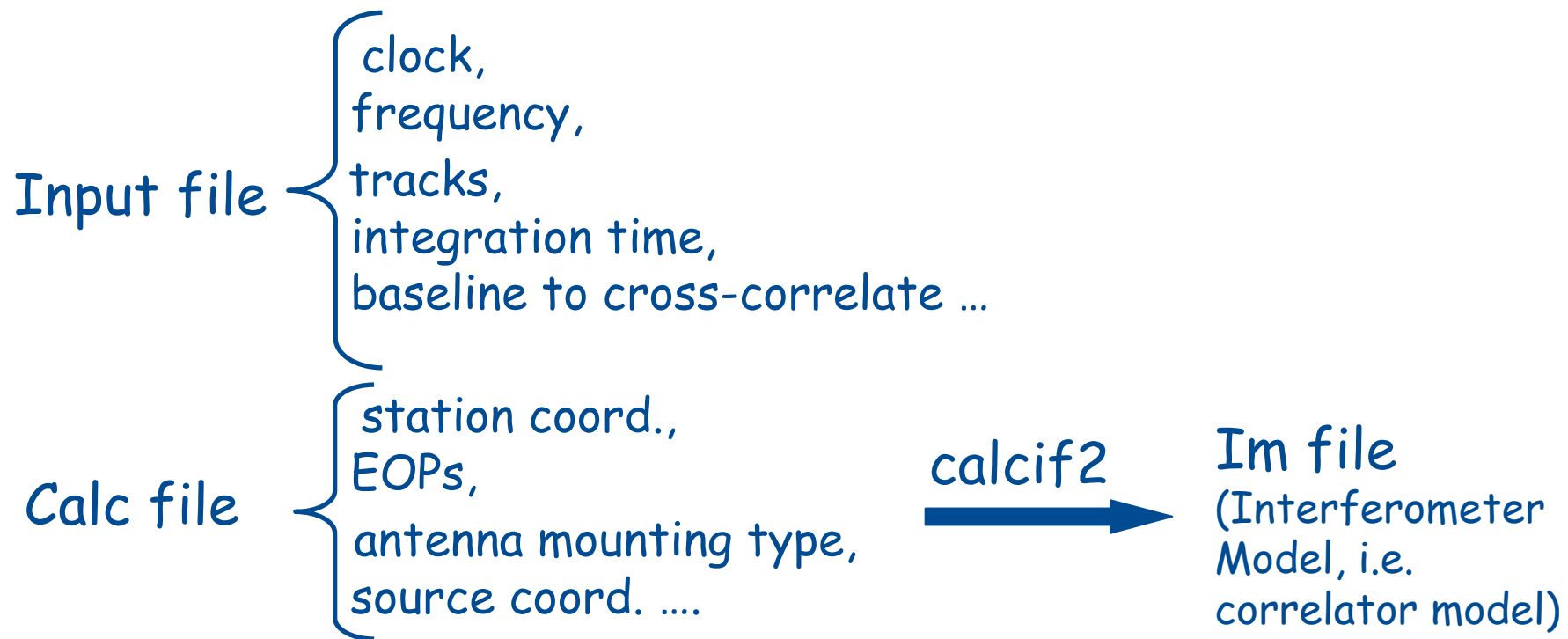
USB

From tables above:

Mk 4	VSI=geo	VSI=astro
1US	0	0
(BBC01)		

VSI output = 0, i.e. TRACK = 02

DiFX requires different files, based on VEX
vex2difx creates those files:



More info:

<http://cira.ivec.org/dokuwiki/doku.php/difx/vex2difx>

`vex2difx` requires a `v2d` file.

Layout of v2d file:

`vex` = vex file name

`antennas` = two letter code of the participating stations
(e.g. `antennas` = WF, ON, WZ ...)

`singleScan` = True/False

SETUP r1600

```
{  
  tInt = integration time in second (e.g. 0.2 s, 1 s ...)  
  doPolar = True/False  
  nChan = no. spectral channels (e.g. 128, 512, 1024)  
}
```

suggested value for trial-
correlation

(max for fourfit)

Layout of v2d file cont.:

```
RULE clock{
  scan = scan name (e.g. 222-1700)
  setup = r1600
}
ANTENNA AB
{
  filelist = ab.filelist
  (or) file = path/filename
}
```

In trial-
correlation
mostly only one
scan

How to create filelist:

Directory2filelist: DiFX program to create filelists

`directory2filelist /path-to-data/ Mode > output.filelist` e.g.:

`directory2filelist /raid1/exp1/ Mark5B 512 8 2 > on.filelist`

format Mbps bbc no. nbit

```
/raid1/exp1/No0001 56054.708345 56054.708935  
/raid1/exp1/No0002 56054.709479 56054.710058
```

MJD

NOTE: Mk5B MJD in filelist is offset by 1000 days → filelist needs editing (use linux editors!).

1) run the program *vex2difx*

```
vex2difx r1600.v2d
```

vex2difx creates the files *.input*, *.calc*,

2) run the correlator using the script *startdifx*:

```
startdifx r1600_1.input or
```

```
calcif2 -a → to create the im file
```

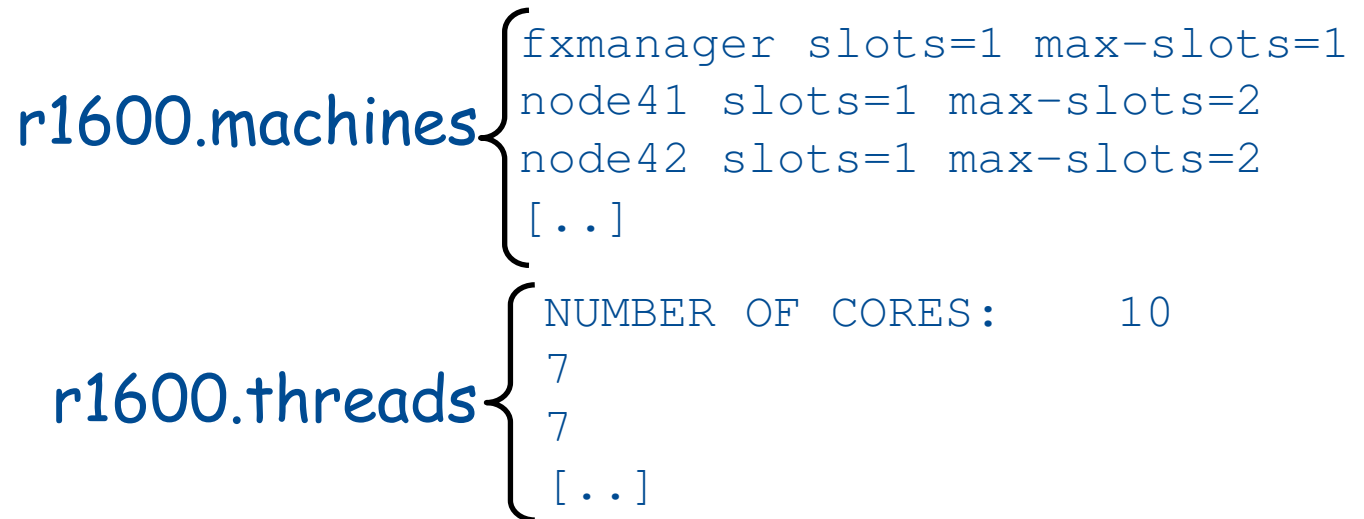
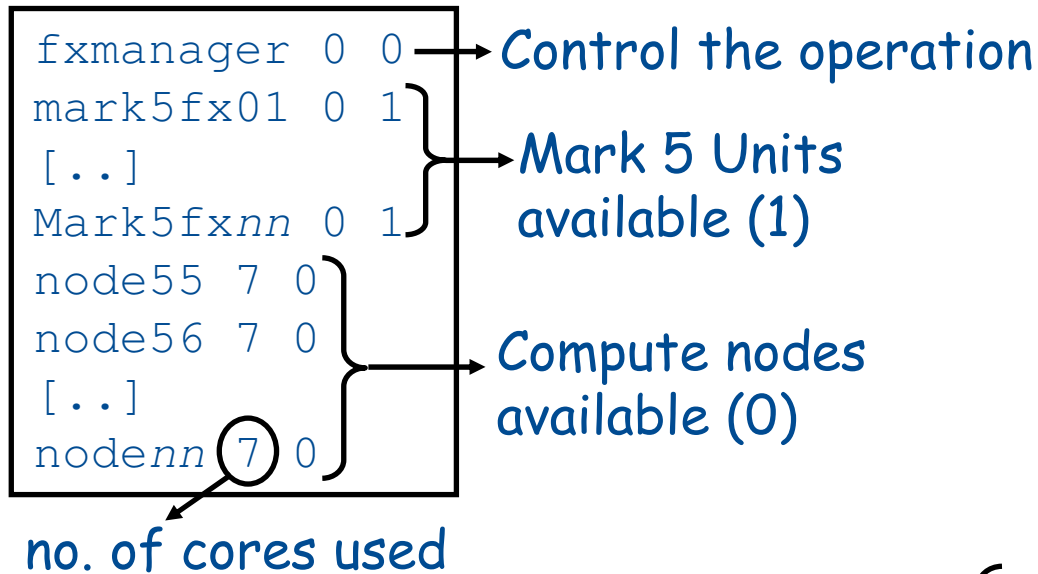
```
genmachines <input file> -> to create the machines and  
thread files
```

```
mpirun -np nn -machinefile <machine file> mpifxcorr <input file>
```

no. of process to start
(found using `wc -l machine file`)

created from *vd2*

The machine file looks like this:



3) *errormon2* shows the correlation details and if all run ...

DiFX creates a directory called `r1600_1.difx` and a file `r1600_1.difxlog`

24

4) run *difx2mark4* to create the files for fourfit:

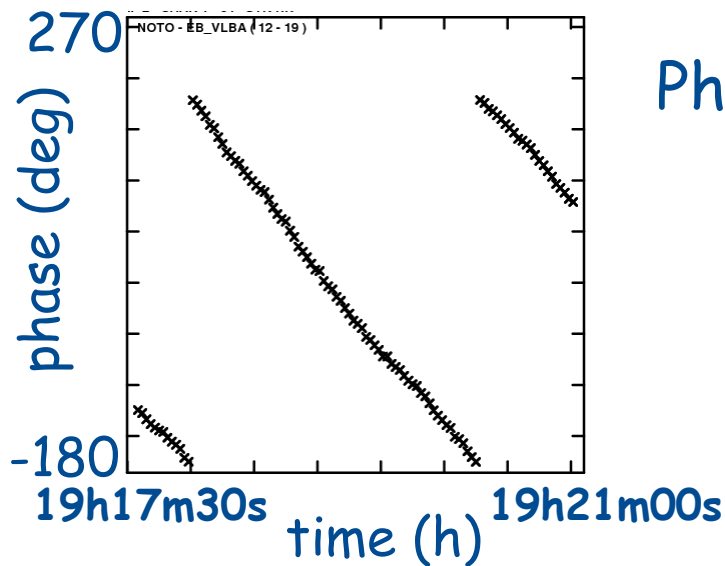
```
difx2mark4 r1600_1.difx (will create a directory 1234)
```

5) run *fourfit*:

```
fourfit -pt -c cf_1234 222-1700
```

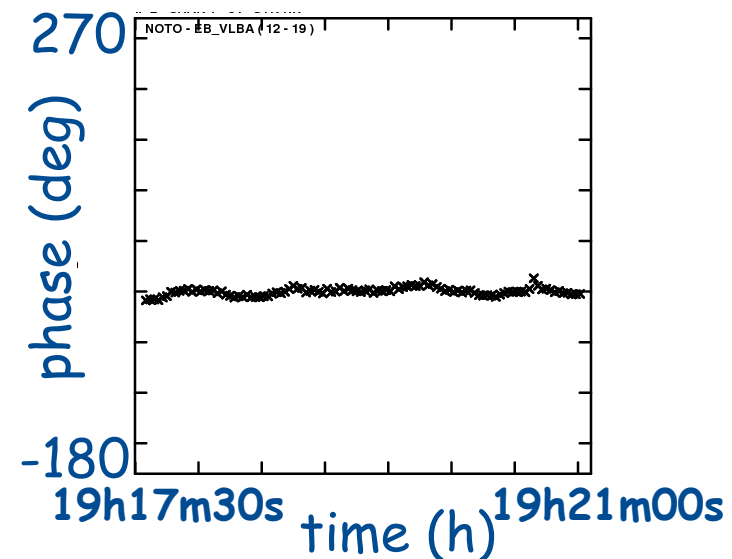
... but why we need to fringe fit the data?

Due to errors in the model, the correlator phases still show a slope vs time:

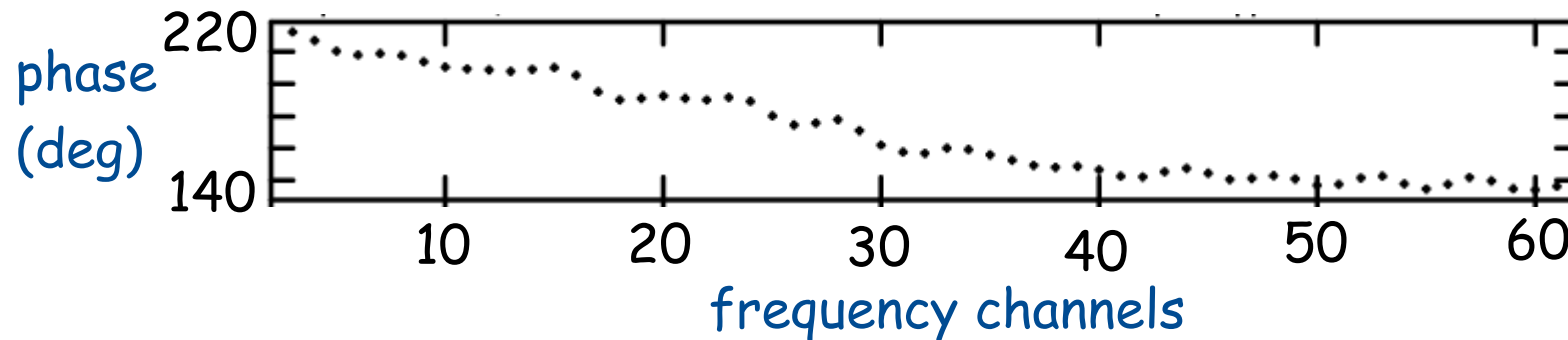


Phase slope vs time is "fringe rate"

Fringe Fit refines the model removing the fringe rate

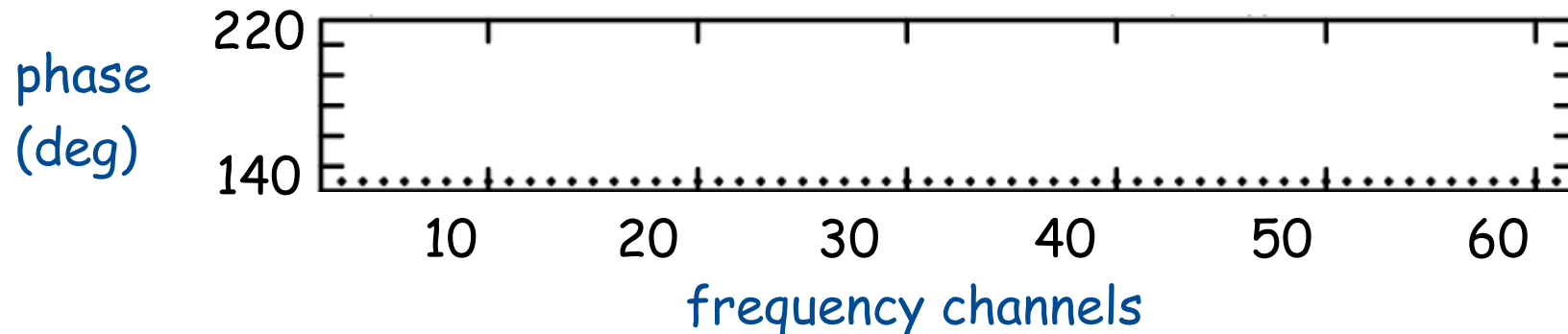


Due to errors in the model, the correlator phases still show a slope vs frequency:

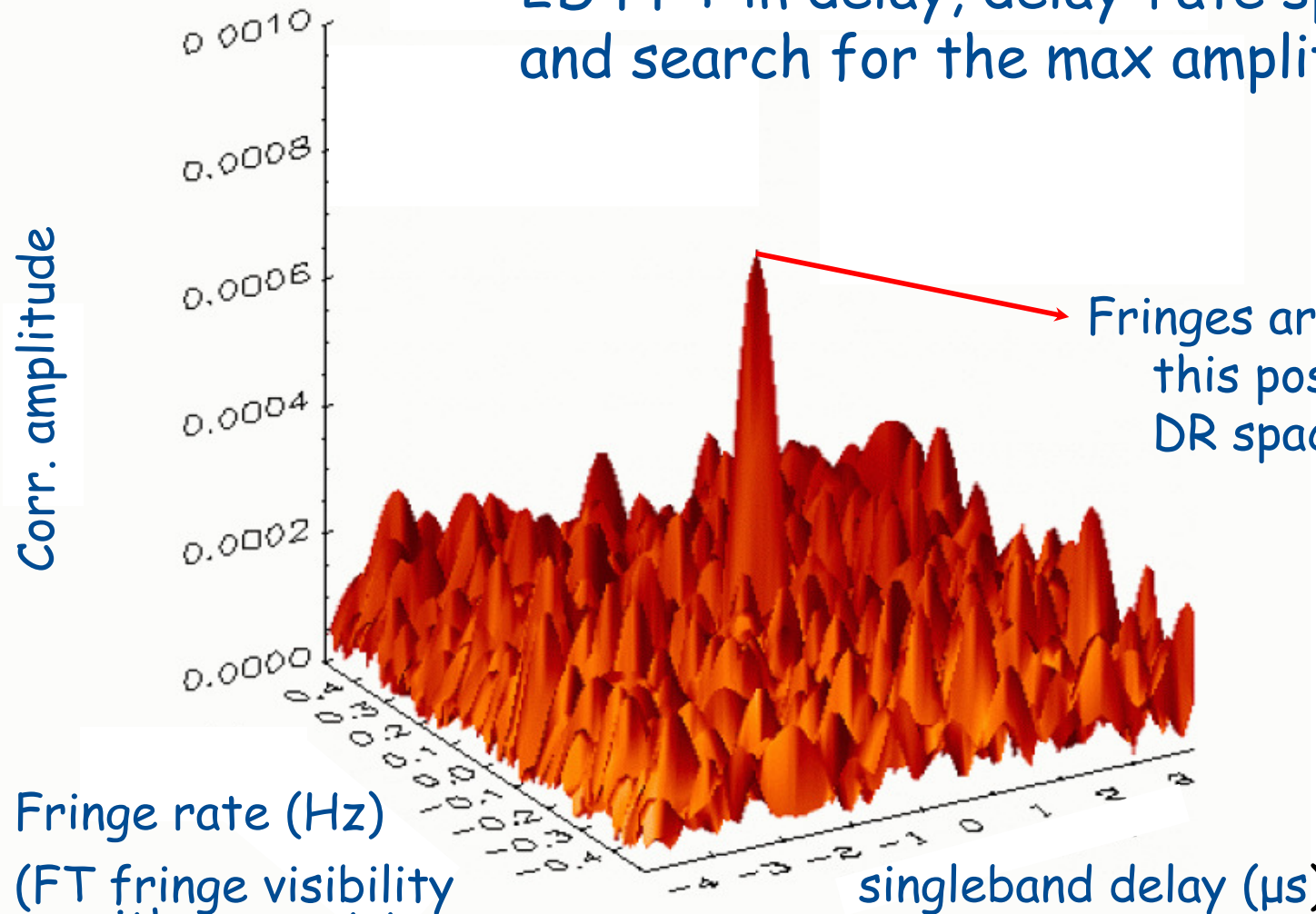


Phase slope in frequency is delay.

Fringe Fit corrects the delay pivoting around a reference frequency



2D FFT in delay, delay-rate space
and search for the max amplitude



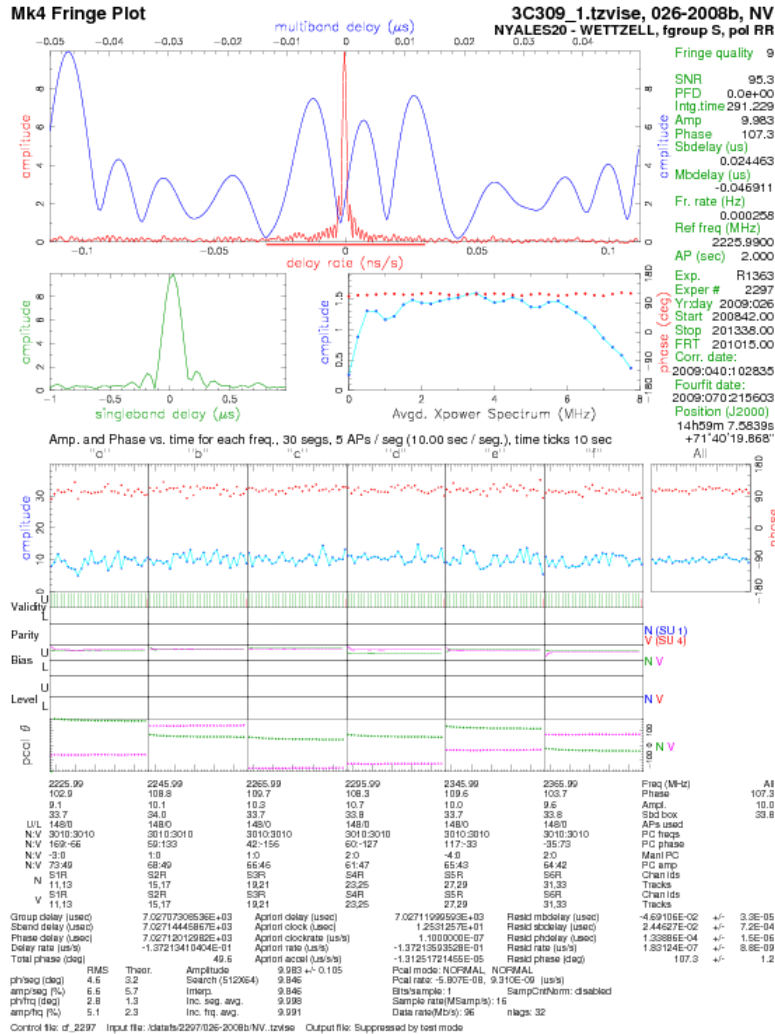
Fringes are located at
this position in SBD,
DR space

Fringe rate (Hz)
(FT fringe visibility
with respect to
time)

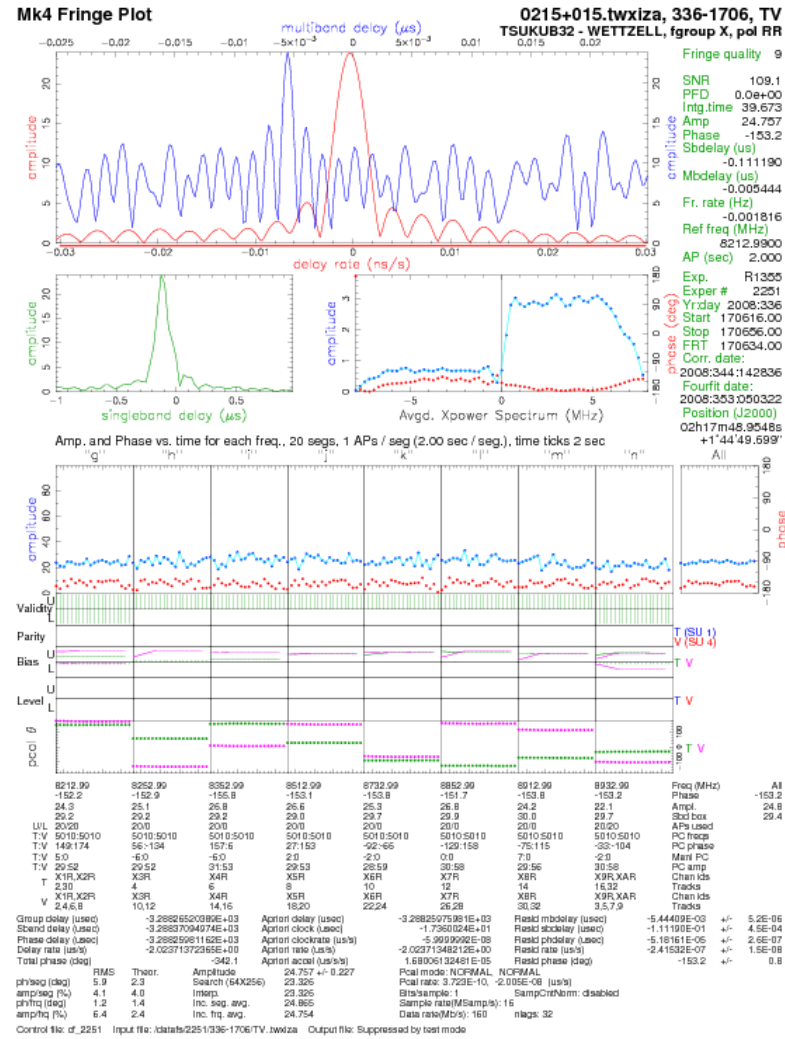
singleband delay (μs)

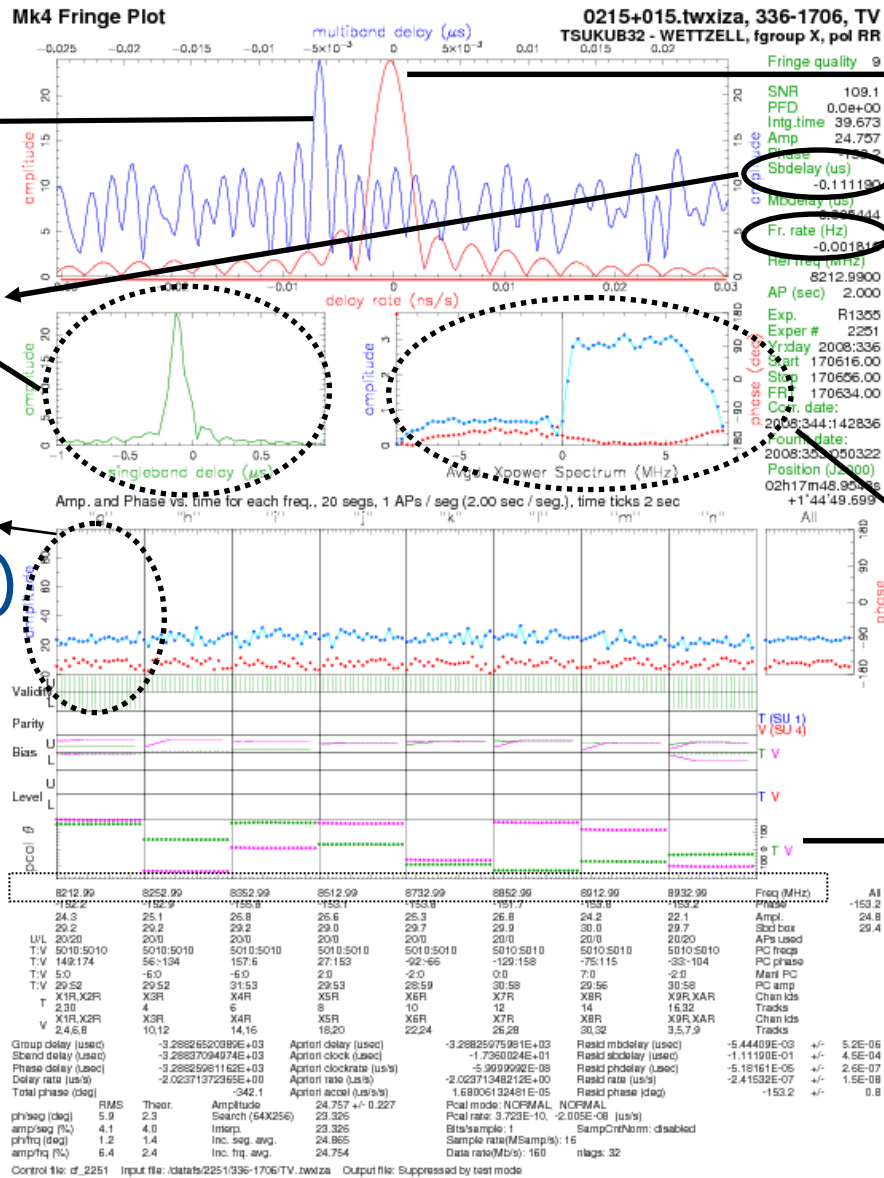
Image courtesy of K. Kingham

S-Band:



X-Band:





Multiband delay (μ s)

Single band delay (μ s)

Phase (red) & amp (blue) vs time for every BBC

Sky freq.

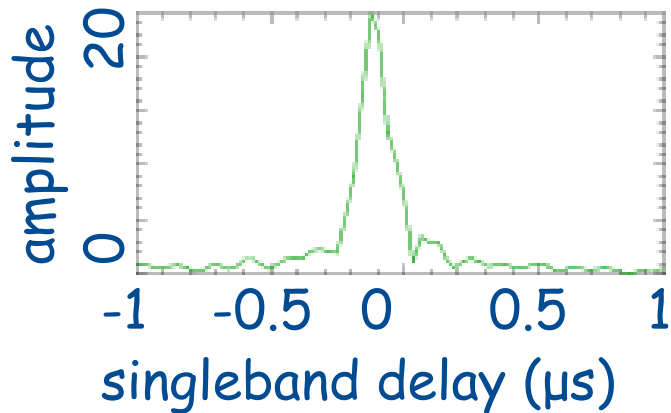
Delay rate.

Fringe rate (Hz) = Delay Rate · Sky freq.

FT of lag spectrum

Pcal phases



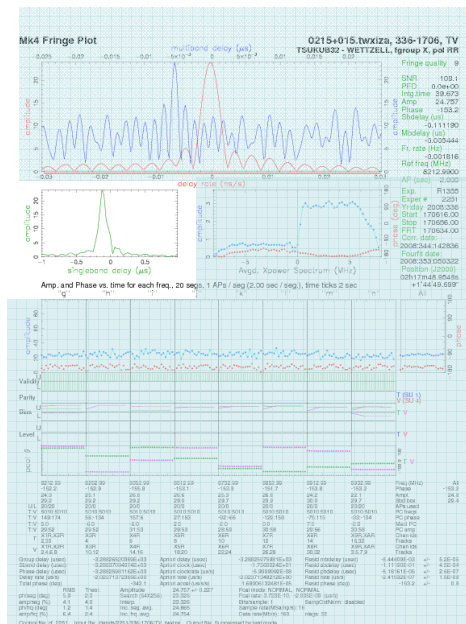


Lag spectrum: output of the correlator integrated over the scan duration.

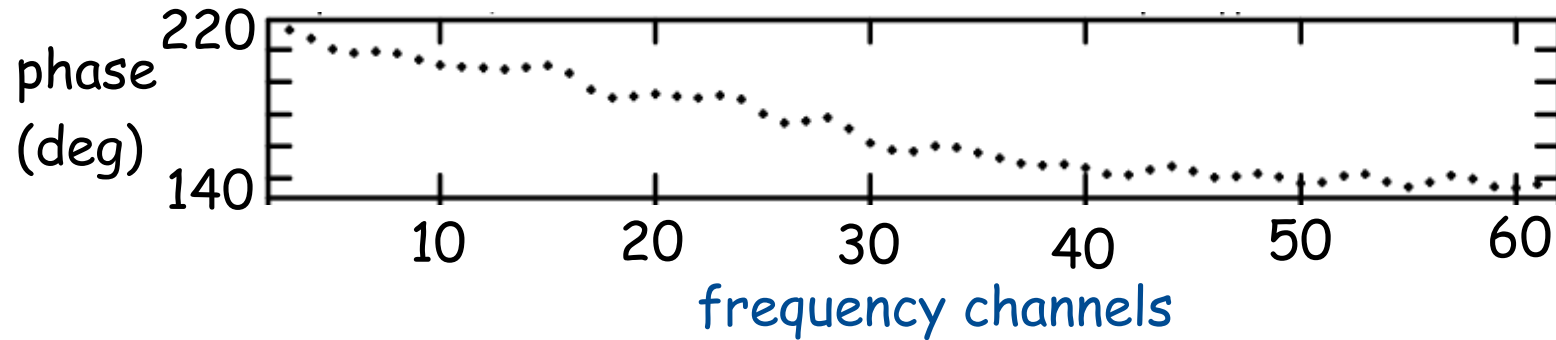
Lag spectrum shown is lag spectra of all BBC stacked.

8 MHz/BBC => 16 Msample/s => sample period = 1 / 16 Msample/s = 0.0625 µs => 0.0625 µs * 32 lags = 2 µs SBD window width.

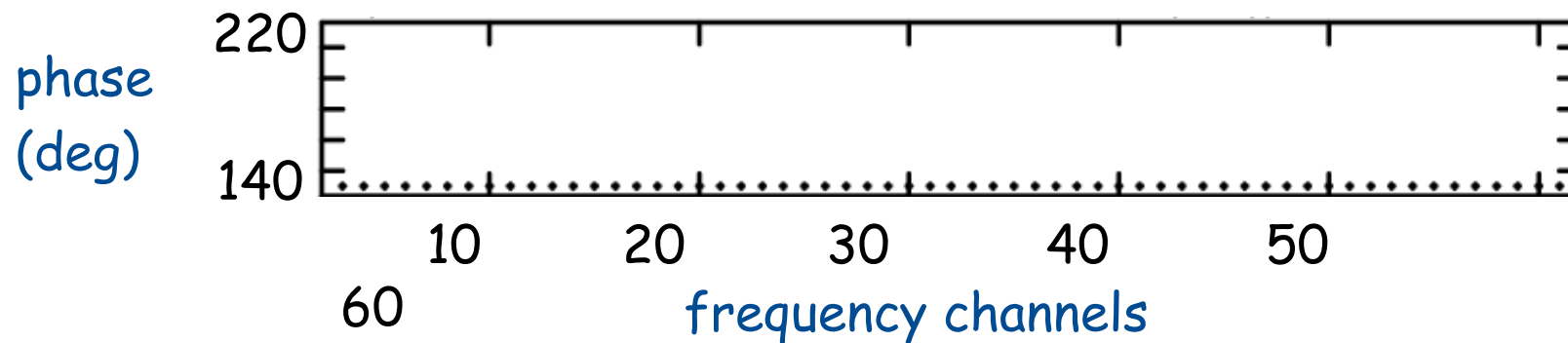
Indicates residual correlator model errors, part of which can be absorbed in the clock offset.

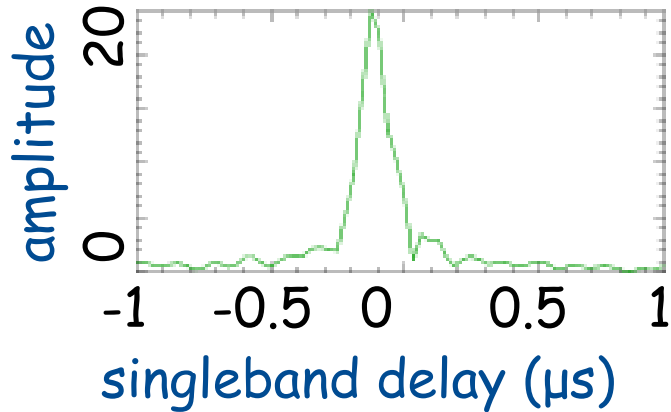


Raw:



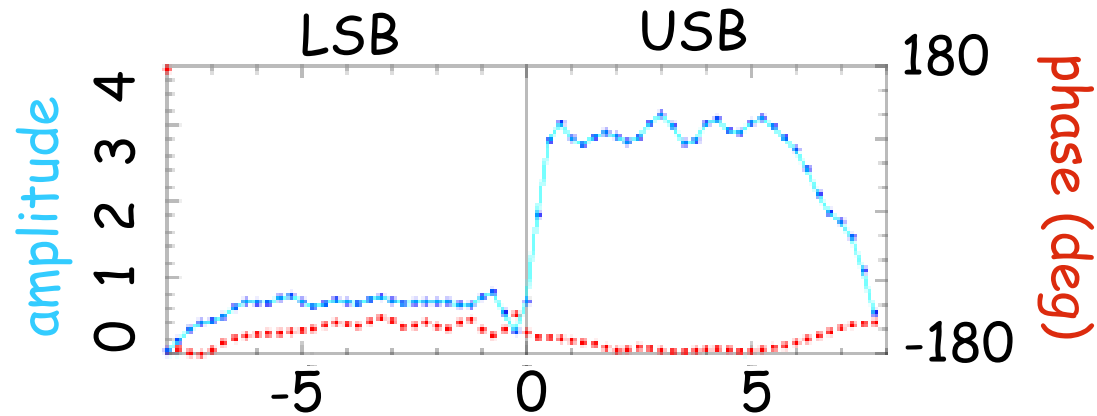
Fringe fitted:



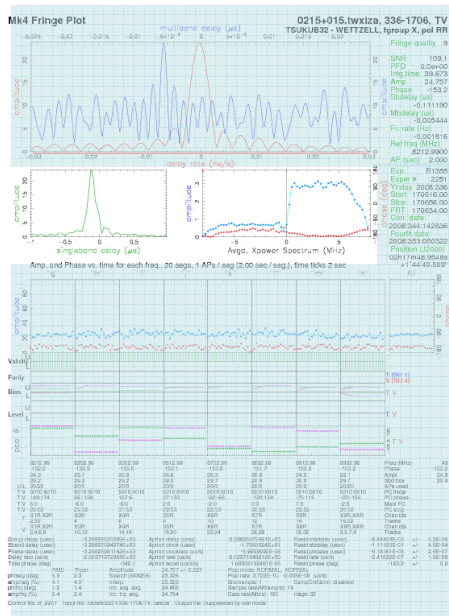


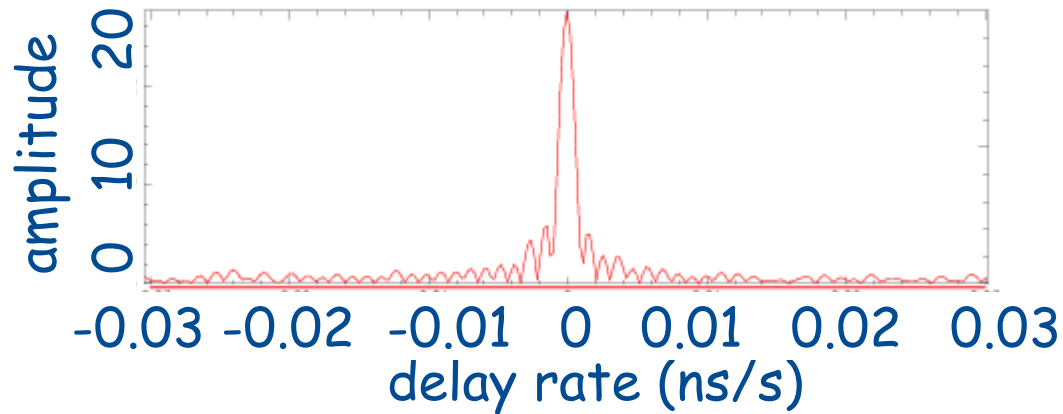
Fourier Transform

$$V(u, v, \tau) = \int V(u, v, \nu) e^{2\pi i \nu \tau} d\nu$$



The data are already fringe fitted.



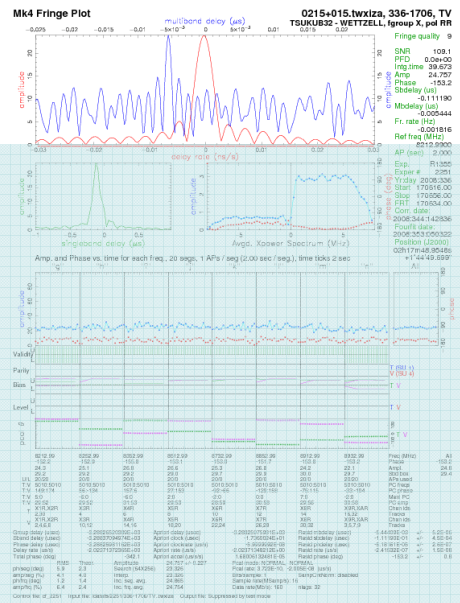


Fringe rate (FR) is the Fourier transform of fringe visibility with respect to time.

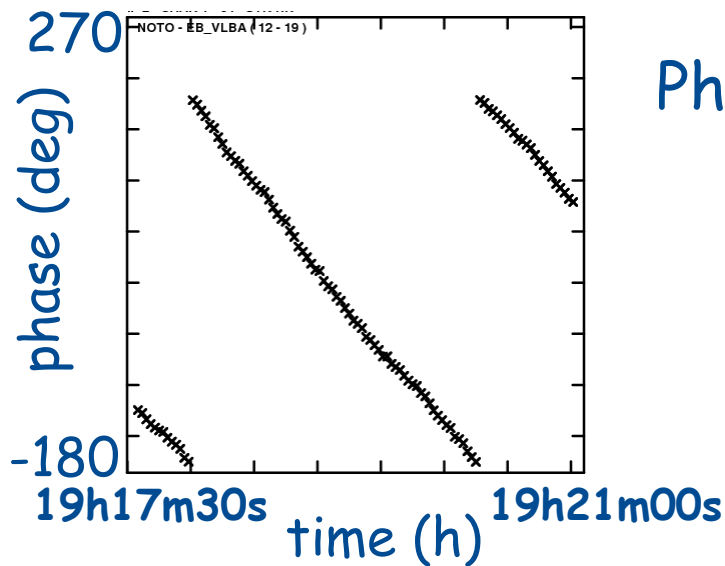
$DR = FR / \text{Observing frequency.}$

$DR \text{ window} = [1 / (2 * AP)] / \text{Obs. Freq.}$

DR tells how fast the fringes move away from the phase centre due to correlator model error. It can be absorbed in the clock rate.

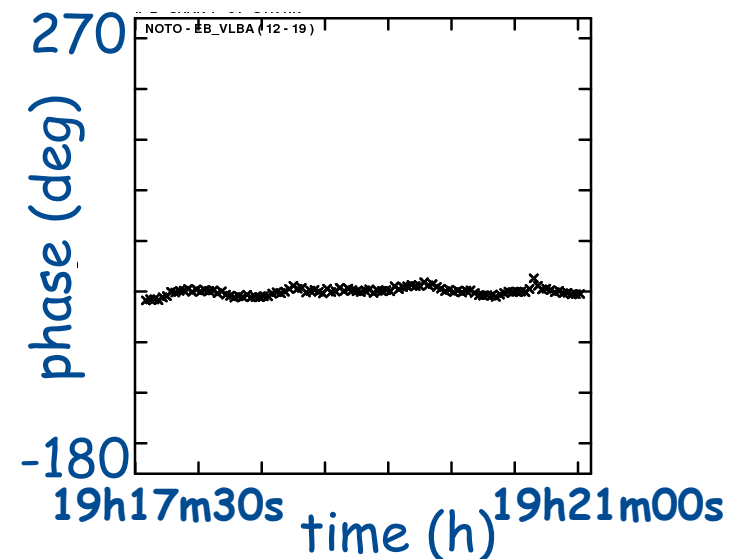


Due to errors in the model, the correlator phases still show a slope vs time:

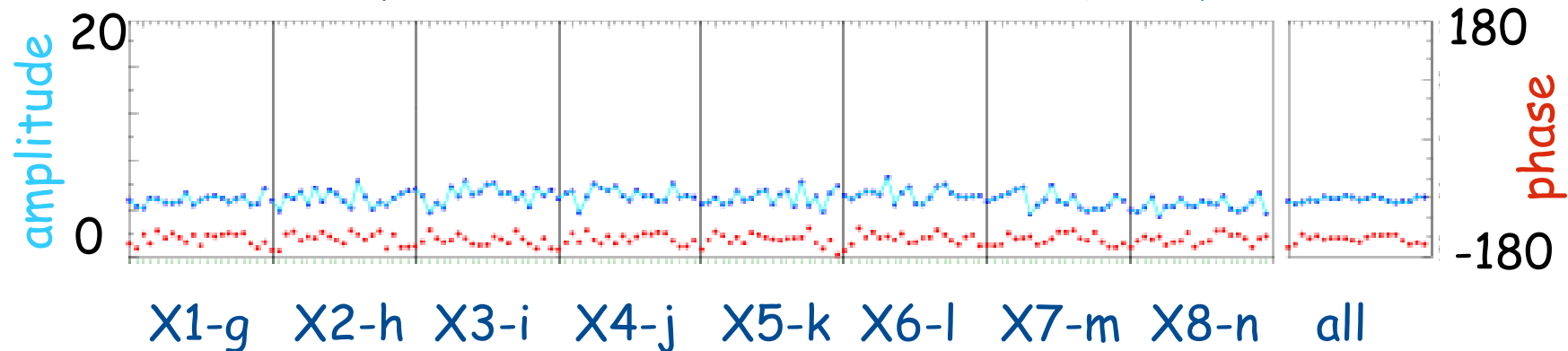


Phase slope vs time is "fringe rate"

Fringe Fit refines the model removing the fringe rate



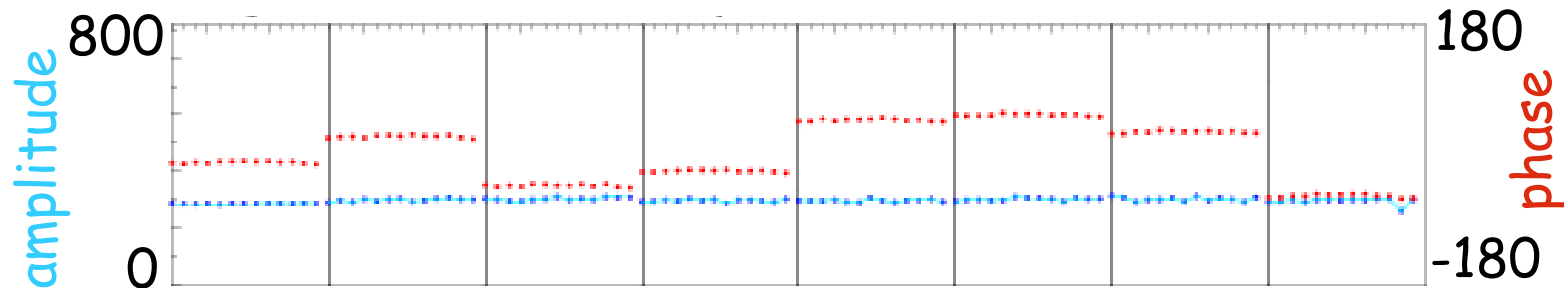
Amp. & Phase vs time for each frequency



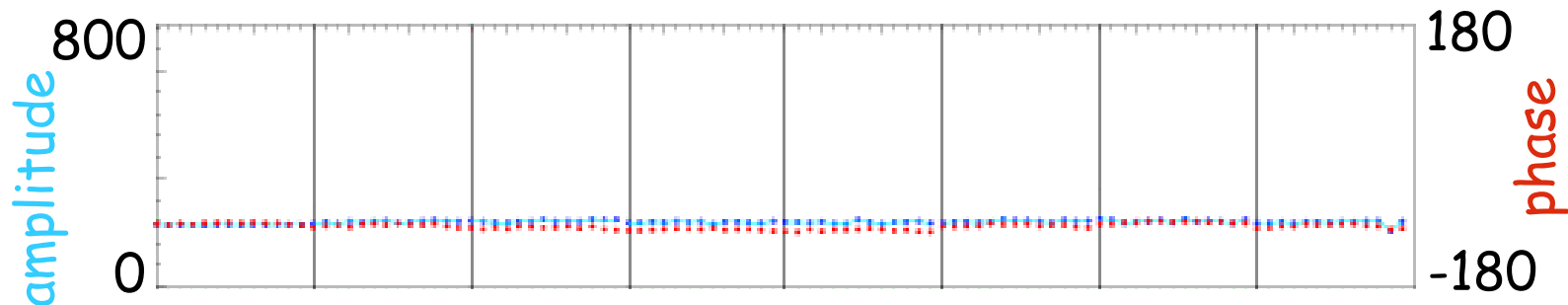
- Every dot represents the phase (red) and amplitude (blue) of the visibility for every segment (~ AP).
- Data are already fringe fitted and pcal has been applied.
- Every BBC/VC channel is represented.

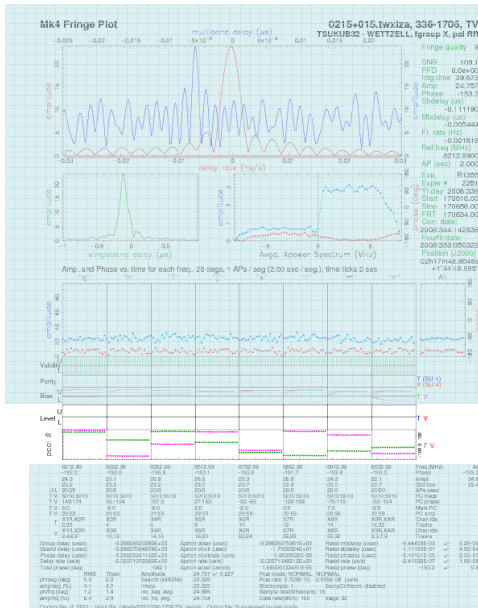
Corrects phase offset of each BBC/VC.

Phase offsets within the BBCs/VCs still present.



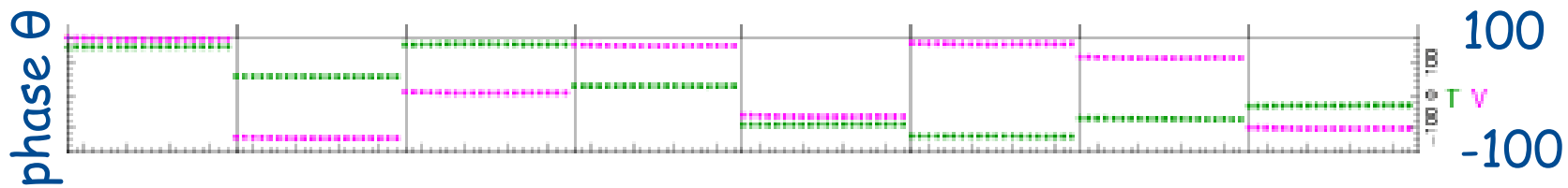
Phase cal phase flattens the phases across the band.





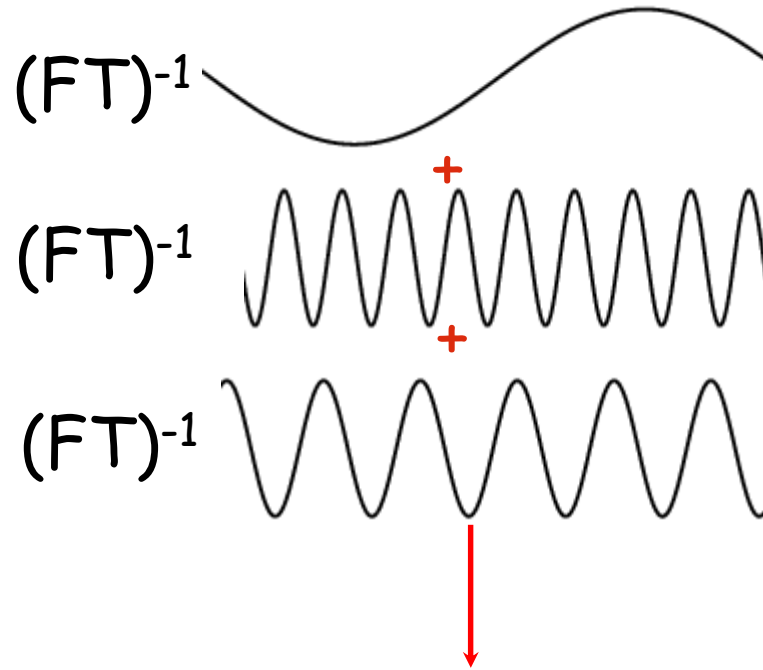
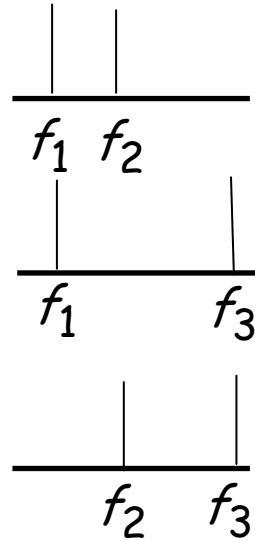
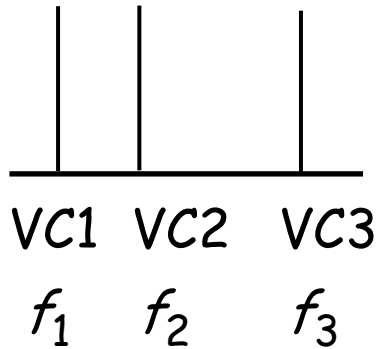
Phase cal phase are plotted whilst only the value of the mean coherent pcal amplitude (PC amp.) is written for each channel.

-
-
-
-
-
-
-
-
-
-

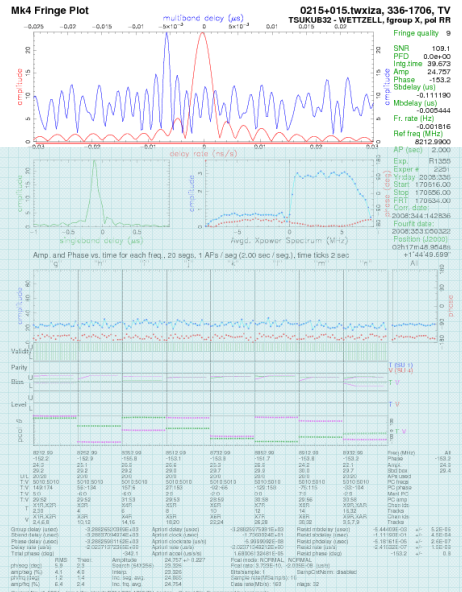
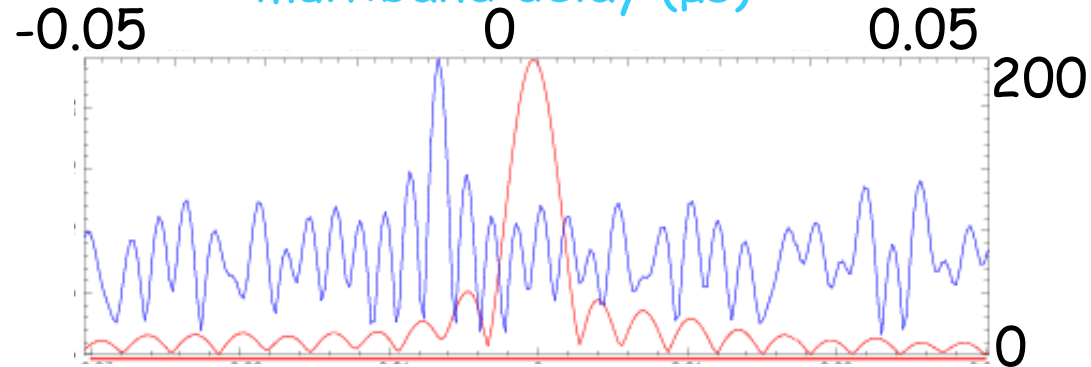


Reference Station (green)

Remote Station (magenta)

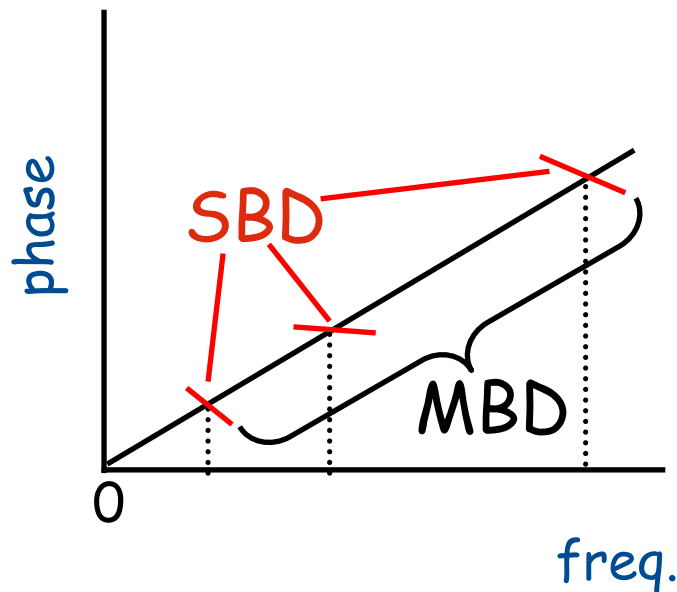


multiband delay (μs)



- SBD = slope of phase across each frequency channels.
- MBD = slope of phase vs whole RF band (e.g. 720 MHz).
- SBD is not corrected by pcal (since fourfit uses only one tone).
- MBD is corrected by pcal.

39



MBD is more precise than the SBD

0215+015.twxiza, 336-1706, TV
TSUKUB32 - WETTZELL, fgroup X, pol RR

SNR =
Peak amp / σ

Ref-Rem, Band, Polarization

Depends on amp. & phase rms
vs frequency and vs time.

Mean visibility amp. & phase

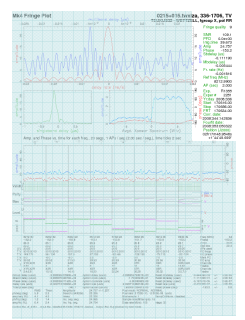
Prob. of false
detection.
i.e. that a
noise spike
exceeds the
signal amp.

```

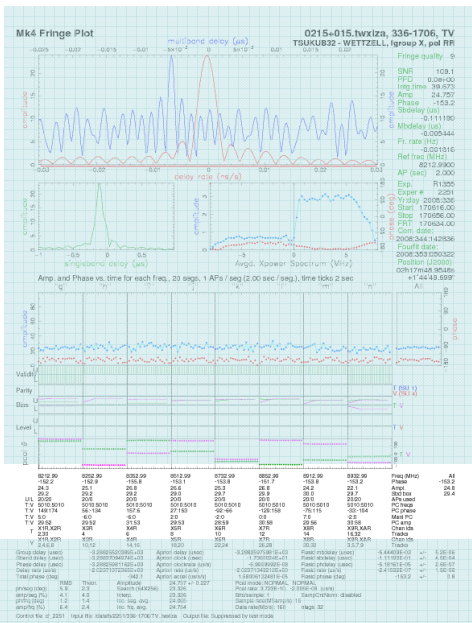
Fringe quality 9
SNR 109.1
PFD 0.0e+00
Intg.time 39.673
Amp 24.757
Phase -153.2
Sbdelay (us) -0.111190
Mbdelay (us) -0.005444
Fr. rate (Hz) -0.001816
Ref freq (MHz) 8212.9900
AP (sec) 2.000
Exp. R1355
Exper # 2251
Yr/day 2008:336
Start 170616.00
Stop 170636.00
FRT 170634.00
Corr. date:
2008:344:142836
Fourfit date:
2008:353:050322
Position (J2000)
02h17m48.9548s
+1°44'49.699"
    
```

Residual SBD (μ s)
Residual MBD (μ s)
Residual FR (Hz)

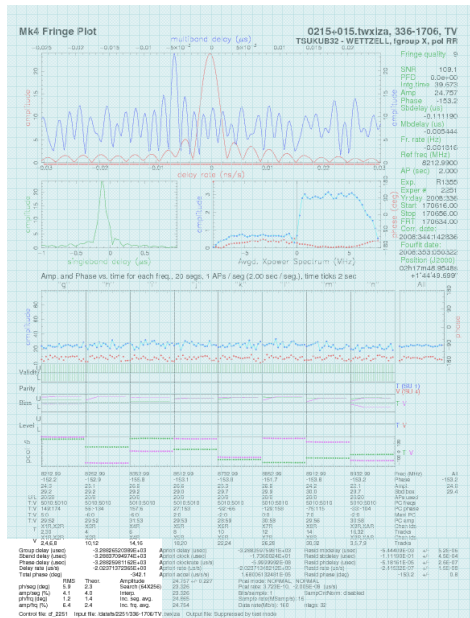
Accumulation Period length,
Fourfit Reference Time, ...



	8210.99	8220.99	8250.99	8570.99	Freq (MHz)	All
	-93.0	-95.7	-99.0	-96.7	Phase	-96.5
	291.1	304.7	308.8	301.2	Ampl.	301.9
	35.6	35.6	35.8	35.7	Sbd box	35.7
U/L	13/13	13/0	13/0	13/13	APs used	
B:N	2010:2010	2010:2010	2010:2010	2010:2010	PC freqs	
B:N	-145:143	-147:30	-14:69	-33:-172	PC phase	
B:N	0:0	0:0	0:0	0:0	ManI PC	
B:N	33:96	33:94	33:93	35:72	PC amp	
B	X1R,X2R	X3R	X4R	X9R,XAR	Chan ids	
	2,4,6,8	10,12	14,16	3,5,7,9	Tracks	
N	X1R,X2R	X3R	X4R	X9R,XAR	Chan ids	
	2,4,6,8	10,12	14,16	3,5,7,9	Tracks	



Correlator model + residual = total



Group delay (usec)	5.61237419104E+03
Sband delay (usec)	5.61251768088E+03
Phase delay (usec)	5.61234964601E+03
Delay rate (us/s)	-4.99305351947E-01

	RMS	Theor.
ph/seg (deg)	1.4	0.3
amp/seg (%)	0.9	0.5
ph/frq (deg)	3.5	0.2
amp/frq (%)	1.9	0.3

Amplitude	301.439 +/- 0.394
Search (32X256)	290.158
Interp.	290.158
Inc. seg. avg.	301.531
Inc. frq. avg.	301.938

rms values of phases & amps. vs frequency: measure of how stable the visibilities are within the total band spanned.



Correlator model applied to the scan

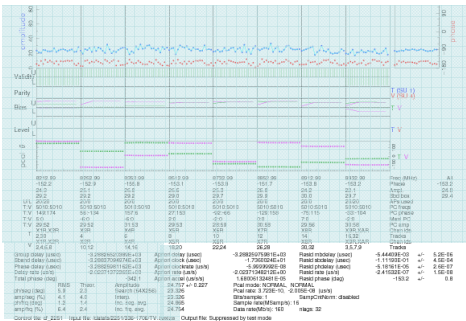


Apriori delay (usec)	5.61234967866E+03	Resid mbdelay (usec)	2.45124E-02	+/-	1.5E-06
Apriori clock (usec)	3.1904583E+00	Resid sbdelay (usec)	1.68002E-01	+/-	1.3E-04
Apriori clockrate (us/s)	3.0000003E-08	Resid phdelay (usec)	-3.26489E-05	+/-	3.6E-08
Apriori rate (us/s)	-4.99305122619E-01	Resid rate (us/s)	-2.29328E-07	+/-	3.4E-09
Apriori accel (us/s/s)	-3.38021266504E-05	Resid phase (deg)	-96.5	+/-	0.1

Pcal mode: NORMAL, NORMAL
 Pcal rate: -3.693E-08, -1.556E-08 (us/s)
 Bits/sample: 1 SampCntNorm: disabled
 Sample rate(MSamp/s): 8 nlags: 32
 Data rate(Mb/s): 80



Residual correlator model errors calculated by fringe fit.



Copyright © 2011, IGG. All rights reserved. Output the Suppress to no mode. Page 22

Fourfit's parameters are controlled through a control file:

- Scan start and stop time offset for the data to be considered valid.
- DR, MBD and SBD search window.
- Lower sideband offset: additive phase between LSB and USB when correlating VLBA data against Mark4 data.
- Phase cal frequency tone to be extracted.
- Phase cal mode: manual or normal or AP by AP.
- Phase cal phases specify a list of phases to be added to the visibility phases in each BBC/VC channel (if phase cal mode is normal).

cf_1234 is fourfit control file.

It tells fourfit what to do.

Basic layout:

```
pc_mode normal (pcal applied)
```

```
sb_win -256.0 256.0 mb_win -2.0 2.0 dr_win -30.e-4 30.e-5
```

sbd search
window bounds
(μ s)

mbd search
window bounds
(μ s)

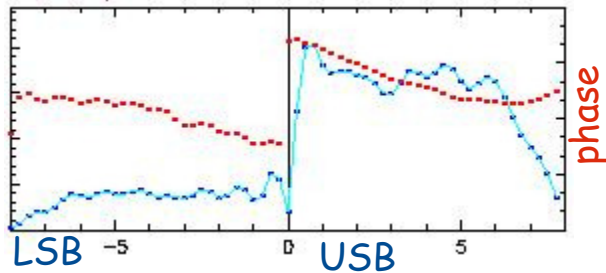
delay rate
search window
bounds

Keep the parameters as above to have a huge window.
If not specified fourfit defaults to a small window !

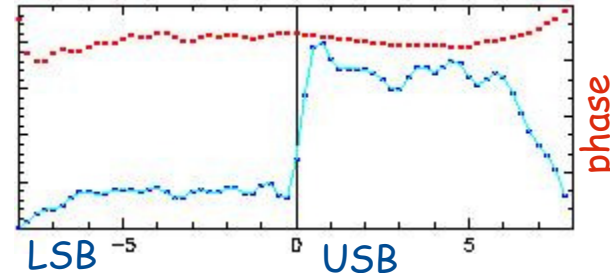
```
if station K
  lsb_offset 260.
```

} LSB/USB offset for different backends

no lsb_offset



with lsb_offset



46

Phase cal tones extracted for X-band in kHz:

```
if f_group X
  ref_freq 8212.99
  pc_freqs ghijklmn 5010 5010 5010 5010 5010 5010 5010 5010
```

→ pivot frequency for fringe fit

Phase cal tones extracted for S-band in kHz:

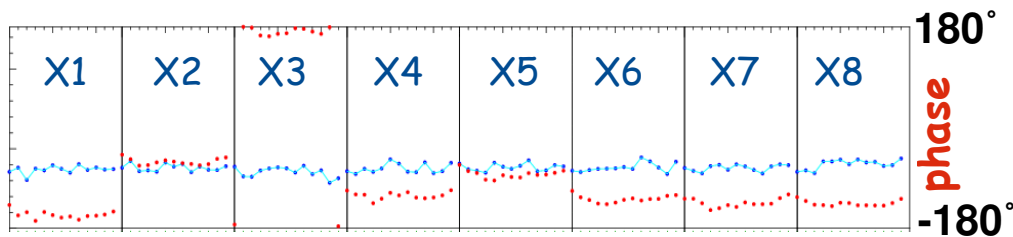
```
if f_group S
  ref_freq 2225.99
  pc_freqs abcdef 3010 3010 3010 3010 3010 3010
```

Manual
phase cal:

```

if station J and f_group S
  pc_mode manual
  pc_phases abcdef -110 -127 -130 -69 -155 -100
if station J and f_group X
  pc_mode manual
  pc_phases ghijklmn 78 123 148 78 115 116 70 104
  
```

47



Manual pcal required!

Additive
phase
(self
cal)

```

if station L and f_group S
  pc_phases abcdef -3.2 0.6 3.6 0.4 0.5 -1.5
if station L and f_group X
  pc_phases ghijklmn -4.0 4.3 4.4 1.1 -0.5 0.8 -6.2 2.0
  
```

Copy raw data (~ 3 MB) onto file and check the data with linux command *od*:

`od -tx4 file name`

output is like:

	frame no.	time stamp:MJD & second of day	fractional second & header error check
0000000	abaddeed bead0001	0974ad5f	f00abf01
0000016	0d645d49	57143f17	3a19c152
			a0ec5b58

od byte no. in file
data
...

ABADDEED => header sync word (every 10016 bytes)

if lots of hex are zeroes -> no input to DBBC

-
-
-
-
-
-
-



Use `mark5access` library (part of DiFX, but should be possible to install them as stand-alone):

`m5d`: decode data (valid for all data kinds that DiFX reads).

`m5test`: decode data headers and data (valid for all data kinds that DiFX reads).

`m5bstate`: state counts summary (valid for all data kinds that DiFX reads).

`m5spec`: forms total power for each baseband channel in the file (never used by me!).

```

m5d /path/file.m5b Mark5B-256-16-1 10 →
Mark5 stream: 0x89e130
stream = File-1/1=/data10/r1/nyalesund/r1538_ny_171-1212a
format = Mark5B-256-16-1 = 2
start mjd/sec = 97 43922.000000000
frame duration = 312500.00 ns
framenum = 0
sample rate = 16000000 Hz
offset = 0
framebytes = 10016 bytes
datasize = 10000 bytes
sample granularity = 1
frame granularity = 1
gframens = 312500
payload offset = 16
read position = 0
data window size = 1048576 bytes
-1  1  1  1  -1  1  -1  -1  -1  -1  1  -1  -1  -1  1  -1
[...]
10 / 10 samples unpacked

```

```
m5test /path/file.m5b Mark5B-256-16-1 →
Mark5 stream: 0x89e130
stream = File-1/1=/data10/r1/nyalesund/r1538_ny_171-1212a
format = Mark5B-256-16-1 = 2
start mjd/sec = 97 43922.000000000
frame duration = 312500.00 ns
framenum = 0
sample rate = 16000000 Hz
offset = 0
framebytes = 10016 bytes
datasize = 10000 bytes
sample granularity = 1
frame granularity = 1
gframens = 312500
payload offset = 16
read position = 0
data window size = 1048576 bytes
frame_num=2 mjd=97 sec=43922 ns=000625000.0 n_valid=2 n_invalid=0
[.]
frame_num=335990 mjd=97 sec=44026 ns=996875000.0 n_valid=335990
1679990000 / 1679990000 samples unpacked
```

m5bstate /path/file.m5b Mark5B-2048-16-2



Ch	--	-	+	++	--	-	+	++	gfact
0	3937	2332	14736	3995	15.7	9.3	58.9	16.0	1.10
1	3921	8576	8552	3951	15.7	34.3	34.2	15.8	1.10
2	3968	8521	8580	3931	15.9	34.1	34.3	15.7	1.10
3	3833	8597	8651	3919	15.3	34.4	34.6	15.7	1.12
4	3857	8573	8628	3942	15.4	34.3	34.5	15.8	1.11
5	3951	8559	8518	3972	15.8	34.2	34.1	15.9	1.10
6	3947	8642	8416	3995	15.8	34.6	33.7	16.0	1.10
7	3991	8543	8525	3941	16.0	34.2	34.1	15.8	1.10
8	3961	8656	8430	3953	15.8	34.6	33.7	15.8	1.10
9	3934	8582	8531	3953	15.7	34.3	34.1	15.8	1.10
10	3896	8651	8615	3838	15.6	34.6	34.5	15.4	1.12
11	3909	8764	8458	3869	15.6	35.1	33.8	15.5	1.11
12	3971	8613	8531	3885	15.9	34.5	34.1	15.5	1.11
13	3988	8561	8370	4081	16.0	34.2	33.5	16.3	1.09
14	3844	8580	8679	3897	15.4	34.3	34.7	15.6	1.12
15	4002	8445	8581	3972	16.0	33.8	34.3	15.9	1.09

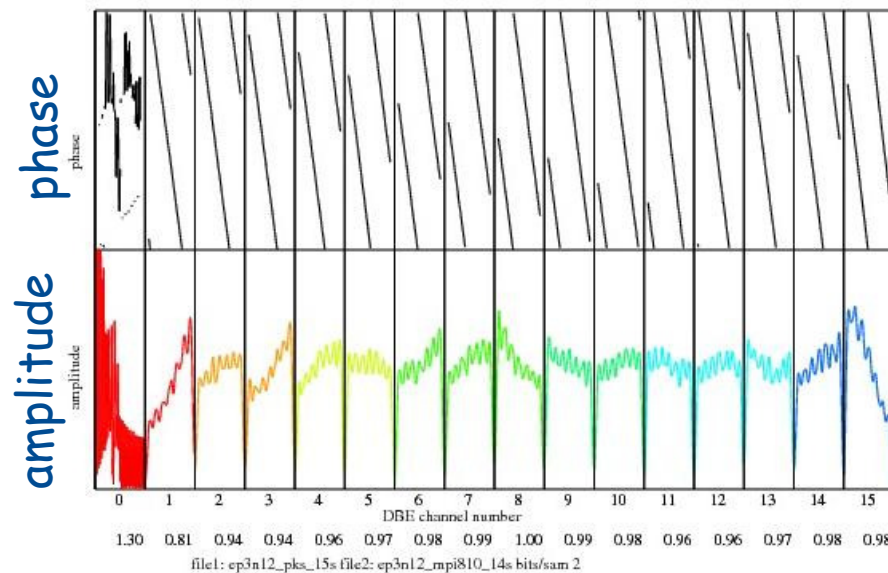
Programs downloadable from Haystack:

`vlbi2` only for 16-channels 2 bit sampling

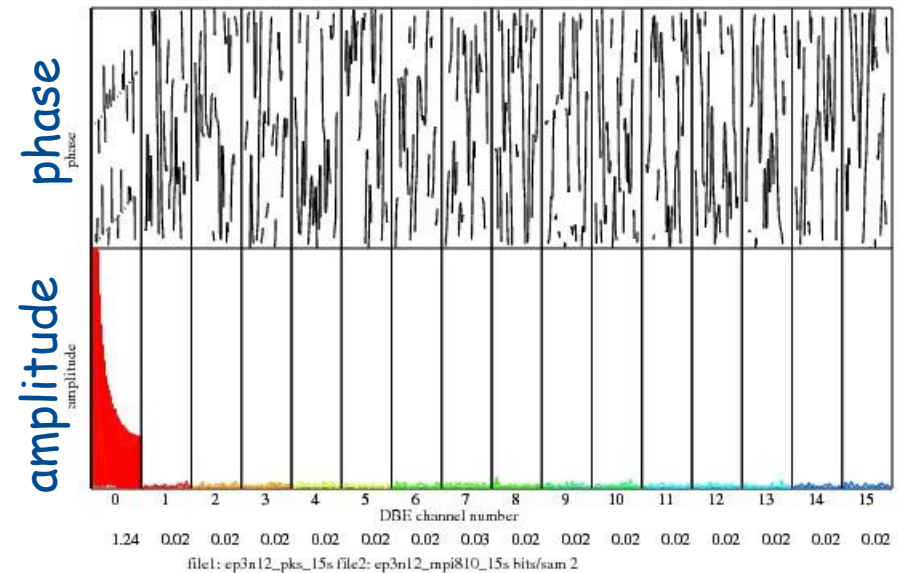
`bstate` only for 16-channels 2 bit sampling

`vlbi0` only for DBE (or equivalent channel assignment)

-
-
-
-
-
-
-

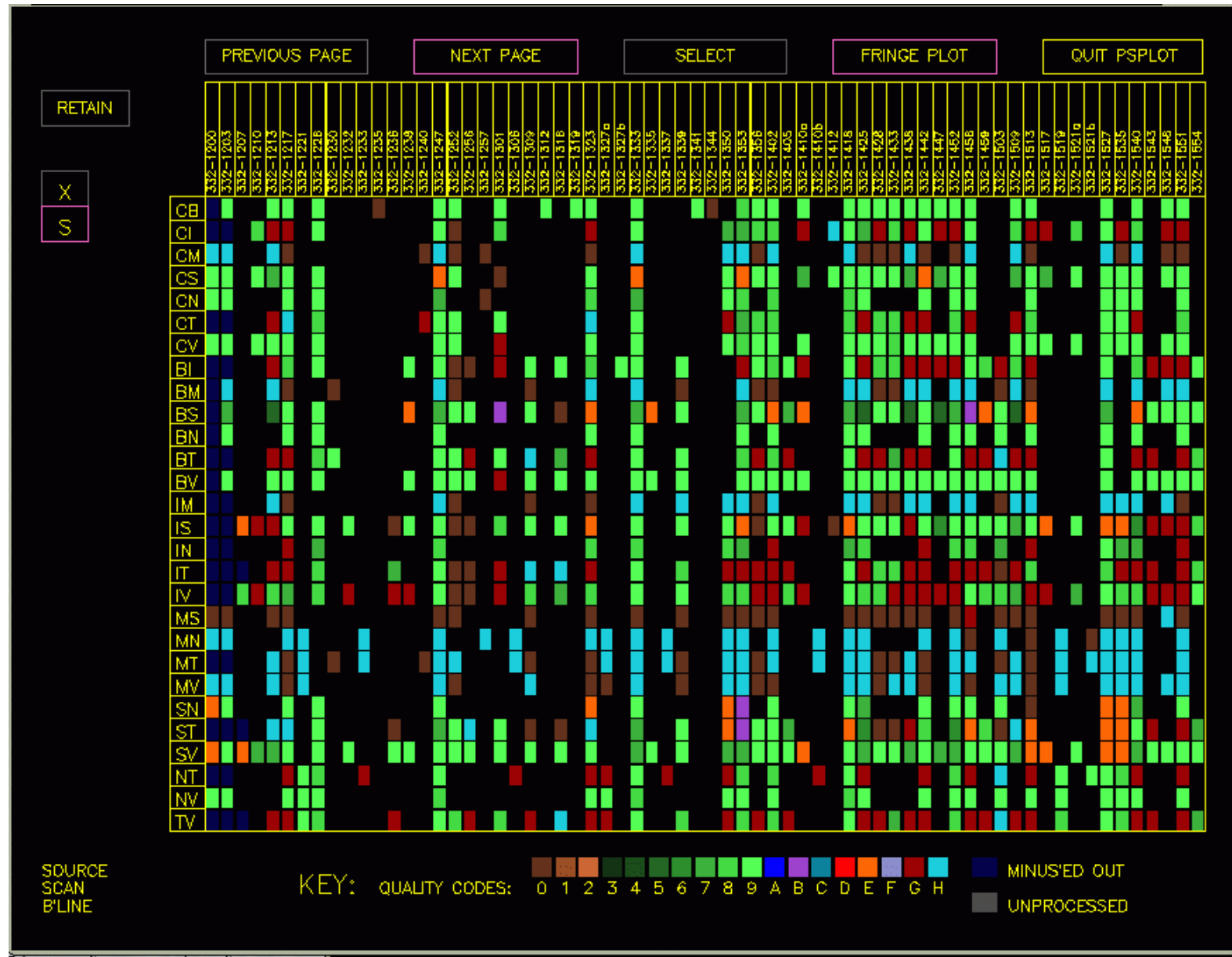


Fringes

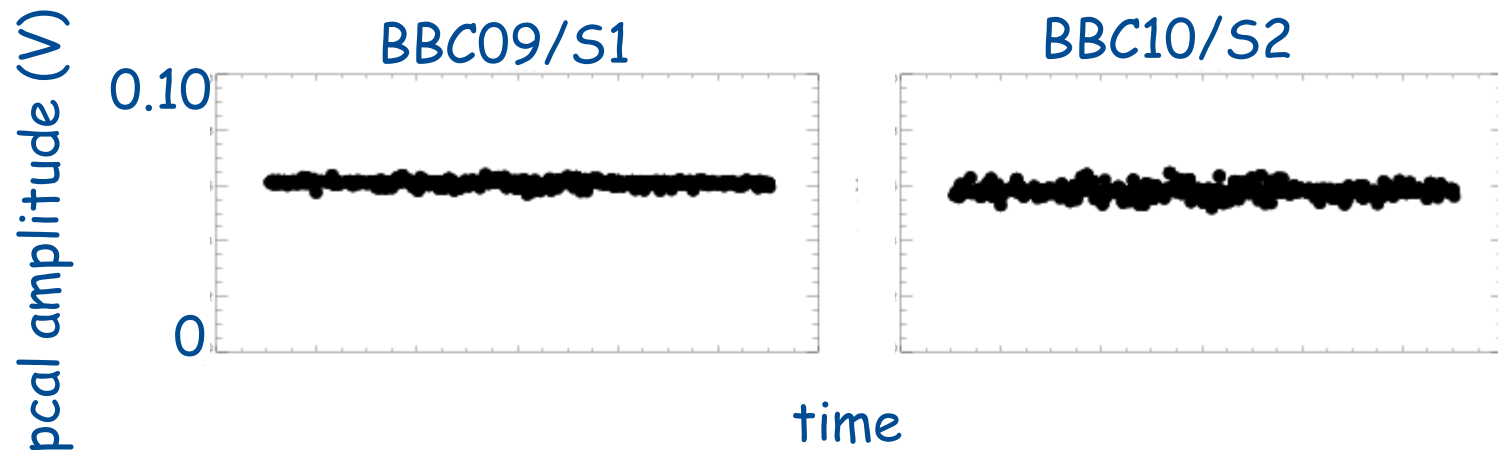


No fringes

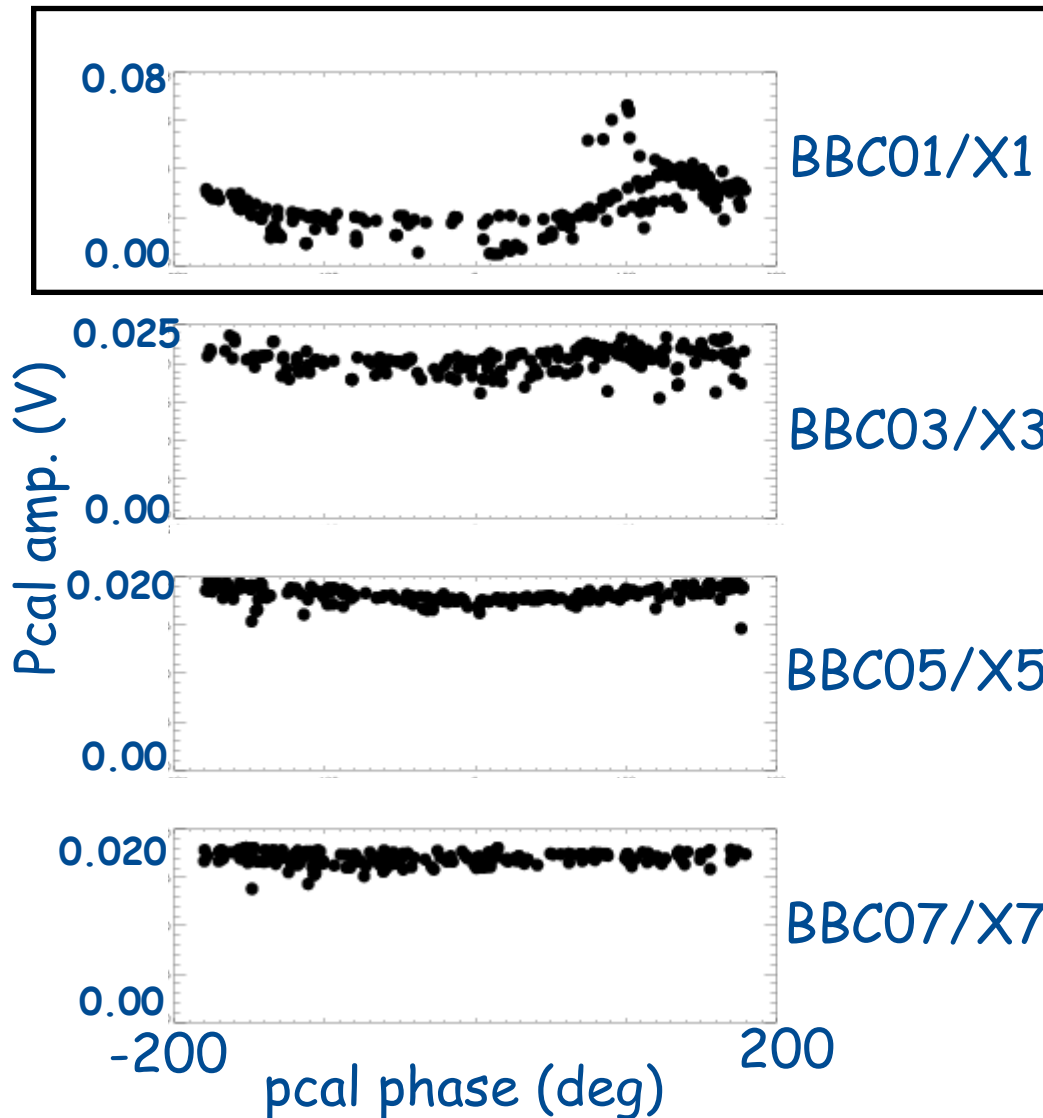
- Import the fringe fitted data
- Check the data quality by plotting
- Check pcal phase and amplitude
- Check SBD
- Check closure quantities for the SBD, MBD and DR
- Export the visibility phases to calculate phase offsets (mostly due to compensate the error between the feed and the pcal injection unit).
- Others... depending in the purpose of the analysis (polarization, source...)



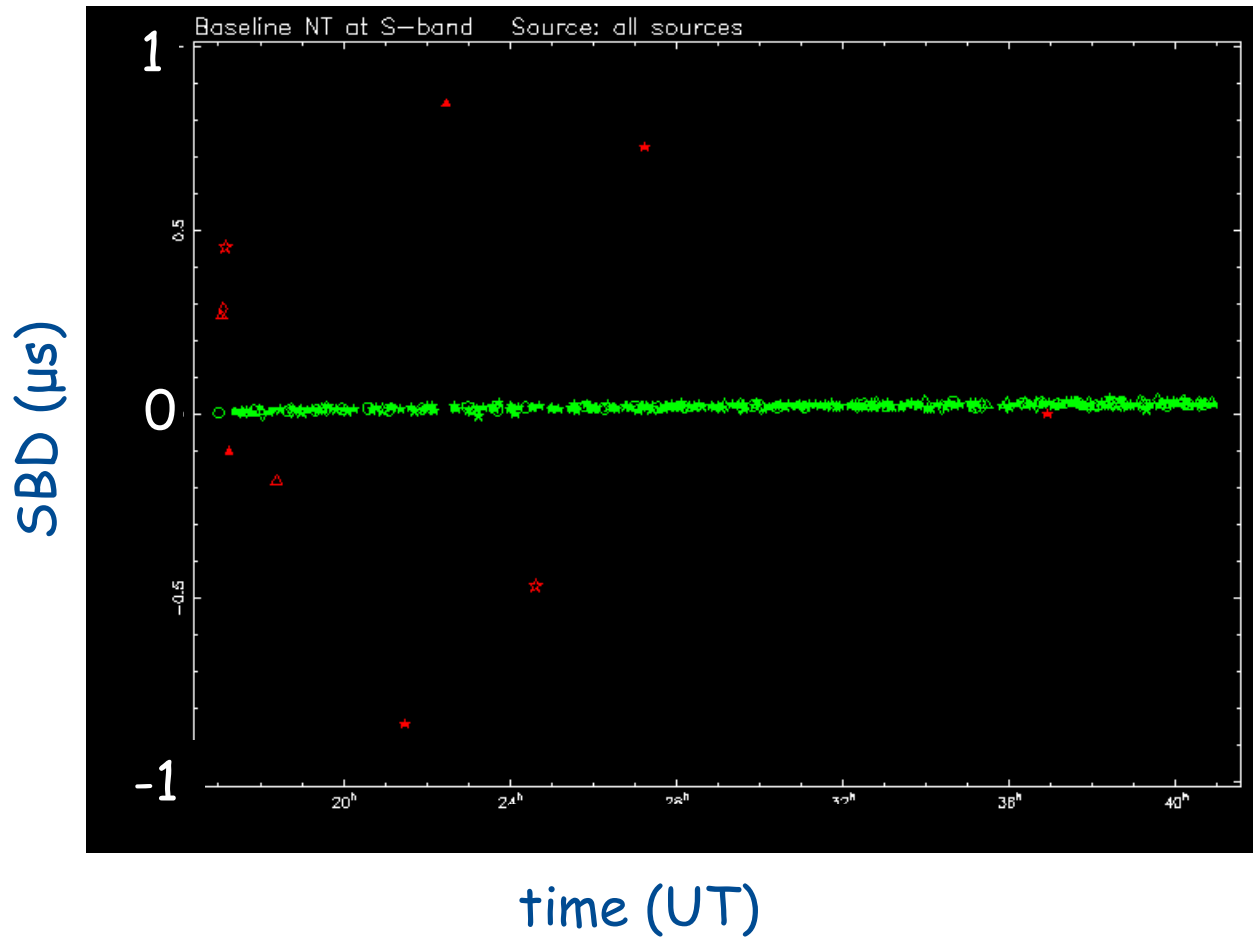
- QC = 0 Fringes not detected ($PFD > 1e-4$).
- = 1-9 Fringes detected, no error condition. Higher number => better quality.
- = B Interpolation error in fourfit.
- = D No data in one or more freq. channels.
- = E Max fringe amplitude at the edge of SBD, MBD or DR window.
- = F "Fork" problem in processing.
- = 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.



- Every station pcal amplitude vs time is checked
- Amplitude variations should be proportional to the inverse square root of T_{sys} . If not, the variation within one BBC/VC or different BBC/VC might indicate a problem: RFI, unlock BBC/VC...

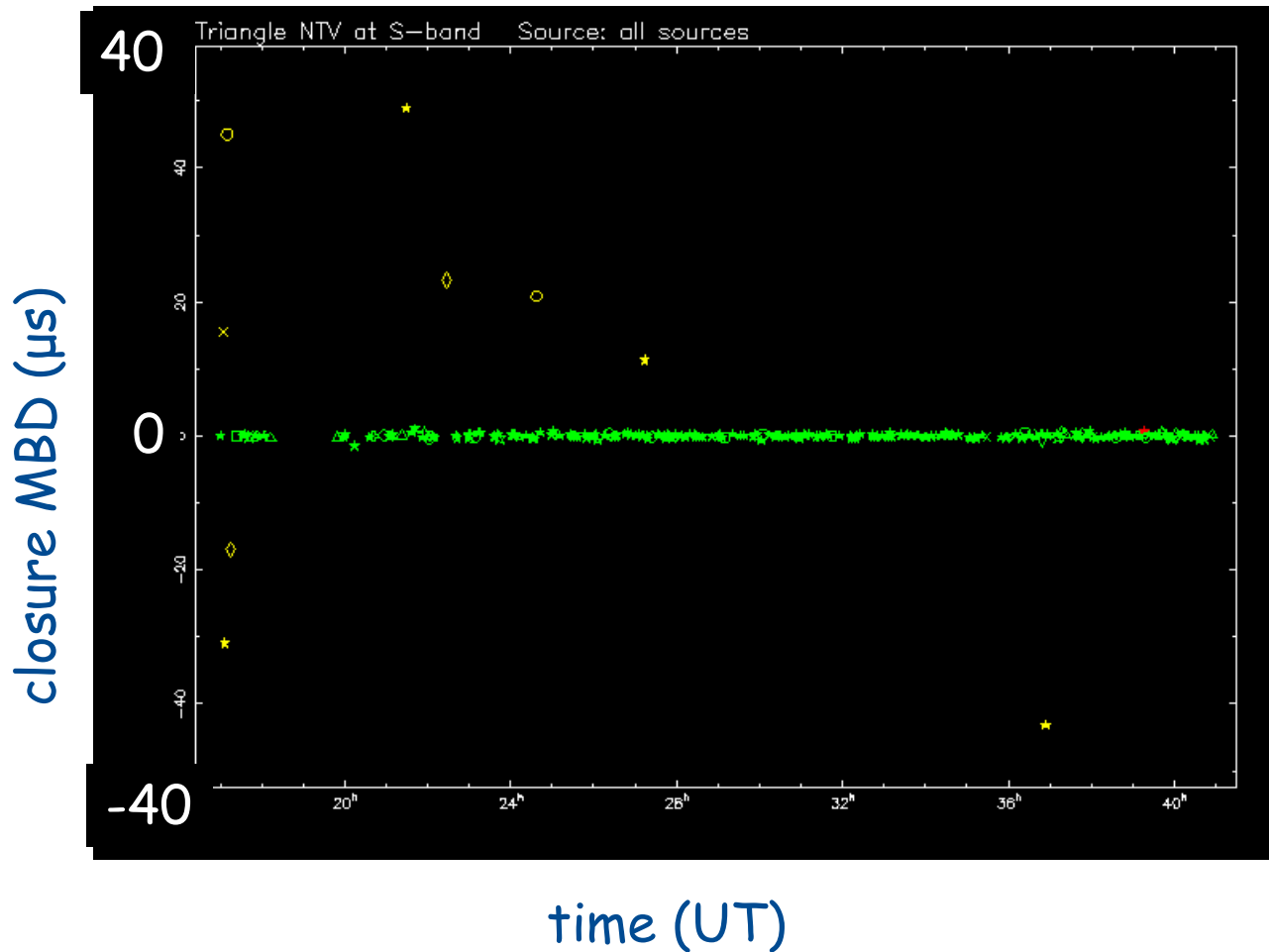


- From these plots we find the spurious signals (sinusoid).
- Spur are narrowband signals coherent with the true pcal and have its same frequency.
- Corrupt the visibility phases.



Check that there are no clock jumps within the observation.

-
-
-
-
-
-
-



Check for station based errors in the data.

- The data are re-fringe fitted using the additive phases and bad channels (e.g. RFI) flagged.
- The data are re-checked using aedit.
- Correlator report is written.
- Stations with problems are notified.
- Database is submitted to analysts.

FX correlator outputs are visibility (real and imaginary components) in the frequency domain.

Lag correlator outputs are correlator coefficients (real and imaginary components) in the time domain.

After correlation, correlator analysts check the data quality (e.g. using *fourfit*).

Sometimes recorrelation is required and performed.

Correlator is a very expensive spectrum analyzer => correlator analysts can help debugging problems at stations.

Correlators deliver to analysts the databases or the FITS file to the astronomers.