EHT/ALMA Correlation with DiFX

Geoffrey B. Crew

MIT Haystack Observatory

mm-VLBI Data Processing Workshop

Leiden– June 10, 2015

Acknowledgements

- A great many individuals have brought us to this point:
  - A. Deller, W. Brisken & the DiFX developer/user community
  - R. Cappallo, M. Titus & Haystack Correlation/Post-processing team
  - APP Collaborators (especially W. Alef, H. Rottmann & N. Pradel for some of the materials presented in this talk).
  - And many others….

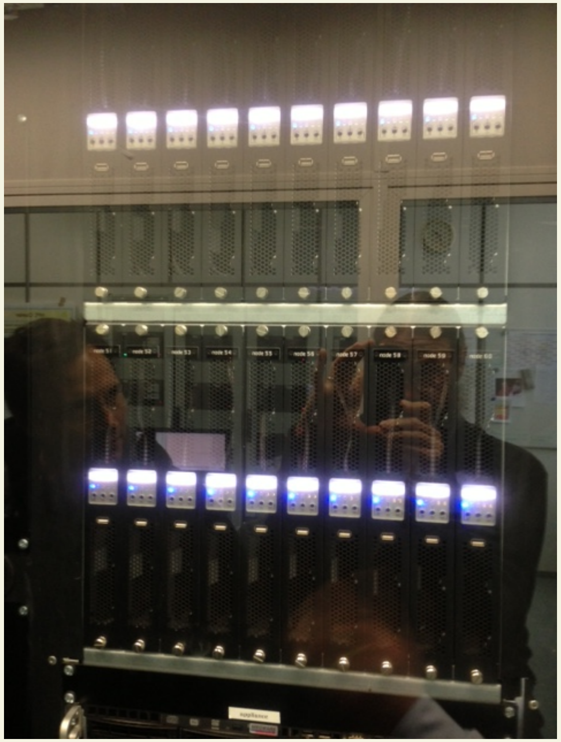
Overview

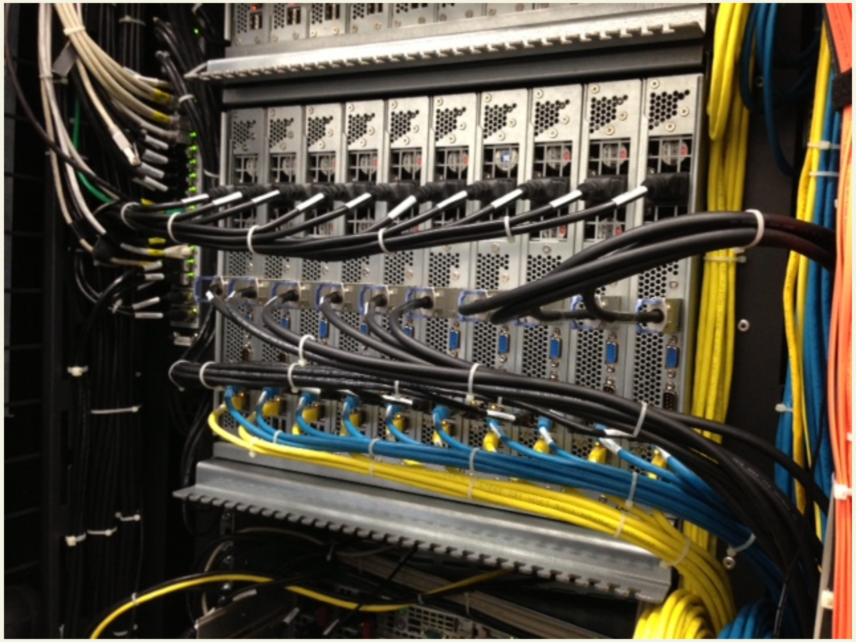
- EHT Correlation Requirements
- Haystack and Bonn Correlators
- Technical issues for DiFX/Post-processing
  - Mark6 playback
  - Multiple data streams
  - Polarization
  - Channels/FFT/DFT/zoom-mode/resampling
  - ALMA sub-band tuning
  - Planned work
- Logistics

EHT Correlation Requirements

- Use DiFX software correlator as baseline plan
- Expansion to ~8 stations over 5-year period
- Bandwidth:
  - Previously 8 Gbps (2 x 512-MHz band, dual pol.)
  - March 2015: 16 Gbps (2-GHz band, dual pol.)
  - Target (ALMA): 64 Gbps (4 x 2-GHz band, dual pol.)
- More than one campaign per year (2? 3?)
- Timely reduction of data (for media reuse)
- ~400 (DiFX) processing cores required

Current Bonn VLBI Korrelator





Cluster Upgrade at MPI

- MPG accepted proposal for new cluster
- ~60 nodes, ~1400 cores, > 40 Gb Infiniband
- > 6 Mark6 recorders (w/expansion chassis)
- > 5x faster than current cluster (depends on Mark6 playback speed, currently unmeasured)
- Procurement for installation Q1/2015 (delayed)
- Performance tuning (since Mark6 is new) will then take several months

Haystack VLBI Correlator(s)

New S/W Correlator

Old S/W Correlator

Cluster Upgrade at Haystack

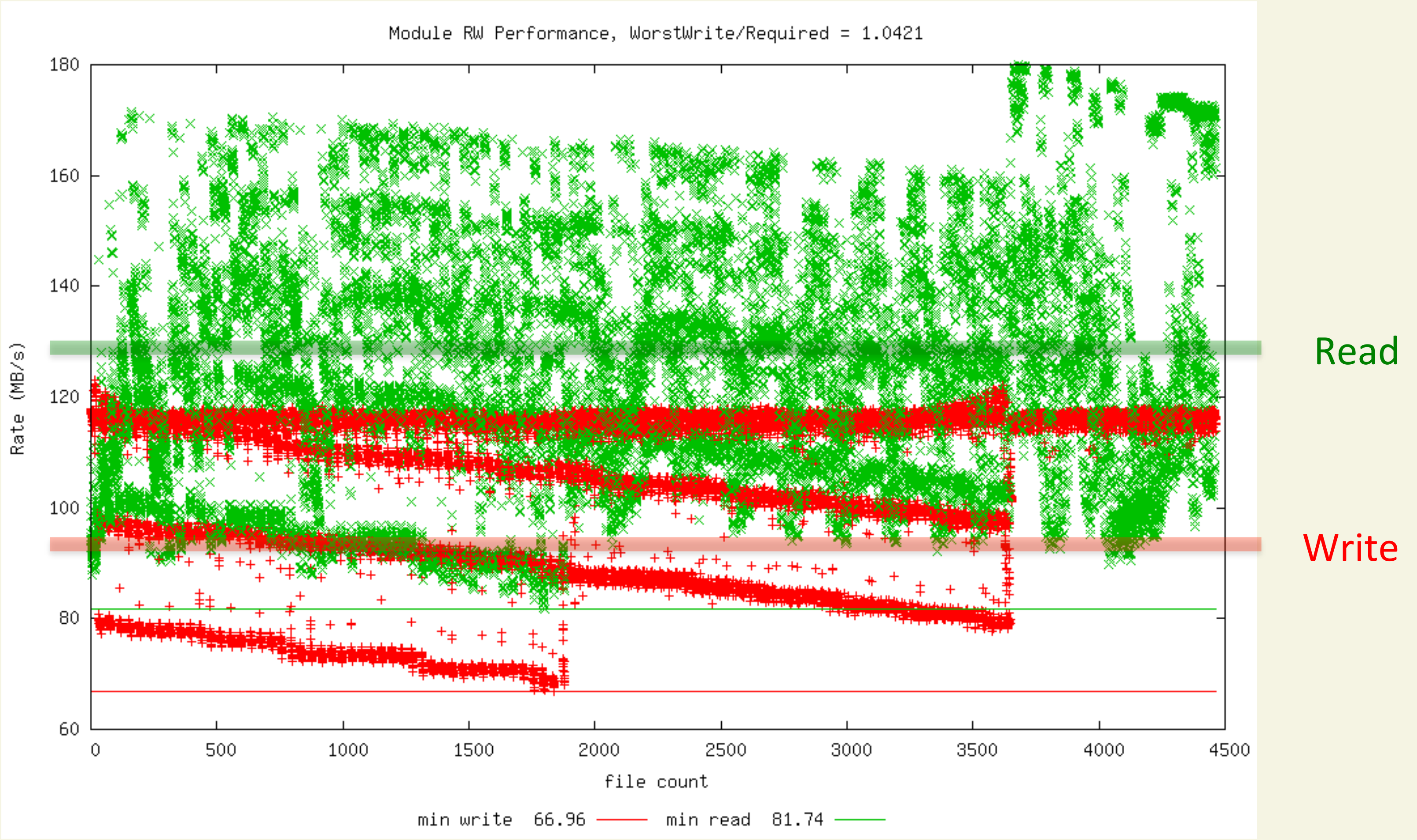
- Phased upgrade strategy:
  - Recent upgrade to support APP commissioning/ March 2015 campaign is in progress (equipment procurement [mostly] complete, installation [around March campaign] in progress)
  - Significant upgrades for March 2017/2018 based on lessons learned from March 2015 production
  - Target is ~8 site playback units with 64 Gbps processed with 4 x 16 Gbps passes

Technical Issues, 1

- Mark 6 Playback
  - 10 GbE limit is 40 Gbps (2 dual-ported NICs)
  - Individual disk limit is ~125 MB/s or ~8 Gbps/mod (required record rate is ~4 Gbps/module)
  - "robust" scatter-gather file system is recorded
  - ***vdifuse*** layer has 10-20% overhead for de-scatter "gather" process. Essential functionality exists now, but robustness & more capability is planned.
  - Mark6 playback into DiFX correlation has been used for ALMA test data so far

Mark6 Disk Testing (8x6TB He)

Read >> Write is good news for playback at correlator



Technical Issues, 2

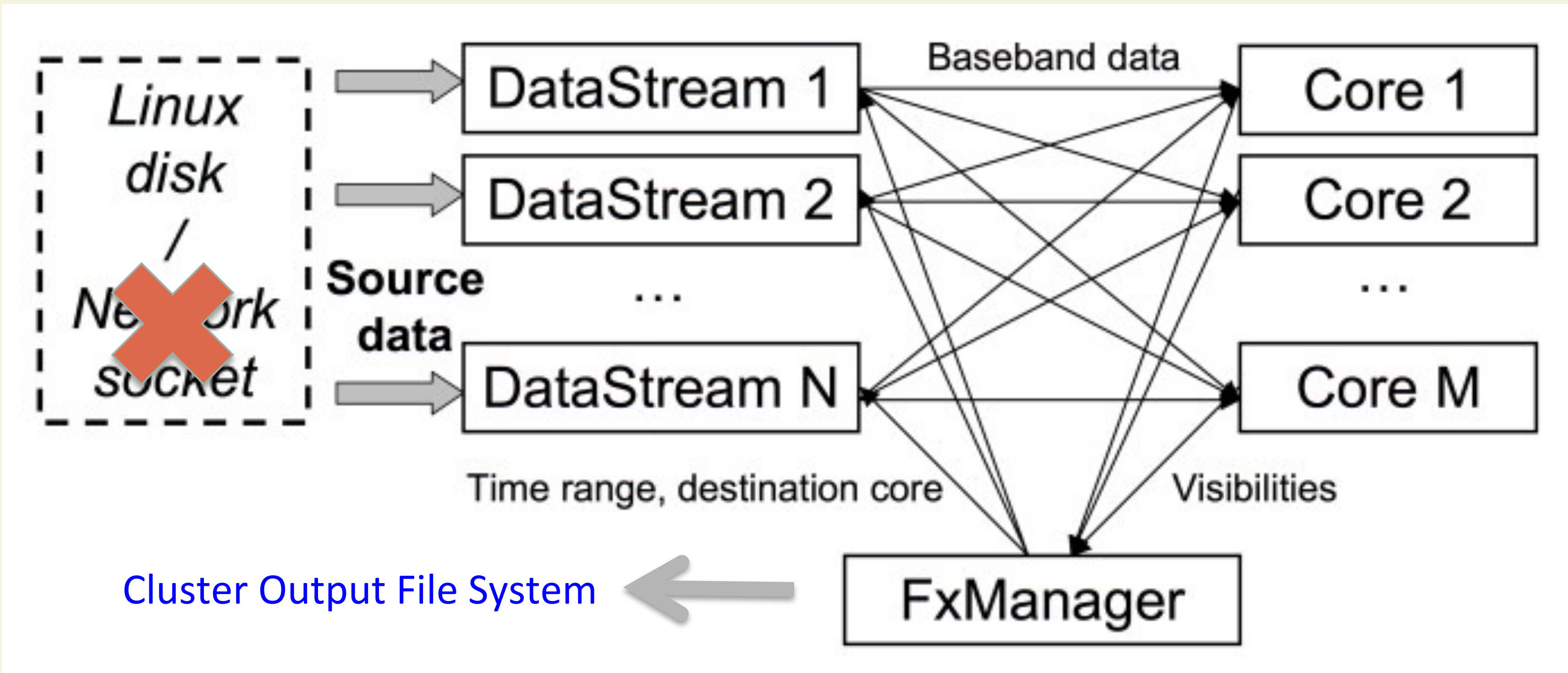
- Multiple data streams
  - (Eventually) 8 x 8 Gbps input streams presented to 4 Mark 6 recorders. Many schemes are possible with VDIF and DiFX; not all are easily supported
  - Recording to groups of 4 modules, but each input stream is onto one 2-module sub-group
  - DiFX ***mpifxcorr*** supports multiple data streams in principal (but not yet [conveniently] in practice)
  - Changes to ***difxio/vex2dif*** are planned (MPI, perhaps in 2015)
  - Changes for the new VEX 2.0, perhaps

DiFX ***mpifxcorr*** Architecture

Mark6 Subgroups

2 Cores/Mark6

Cluster CPUs
(Spare Mark6 CPUs)



At present, we typically assign multiple antenna names per real antenna
And then clean up the mess post-correlation in the analysis phase

Technical Issues, 3

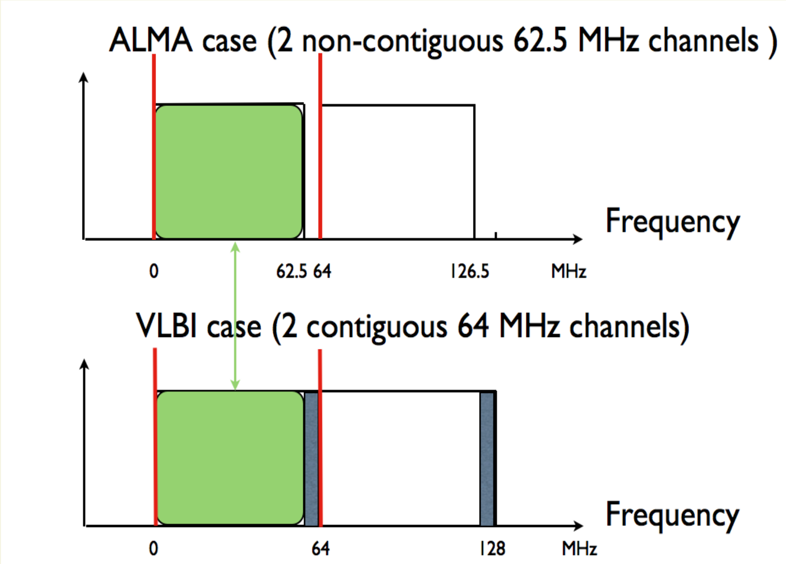
- Polarization
  - ALMA is X/Y, not R/L (quarter-wave plates for 66 antennas was not an allowed option)
  - Lie: DiFX doesn't *need* to know it is X/Y, not R/L in order to construct valid correlation products
  - ***polconvert*** tool developed to "clean up" the DiFX mess (all that lying comes to no good end)
  - Calibration data (for *D* terms) needs to be taken
  - For details see Ivan Marti-Vidal's talk, Wednesday

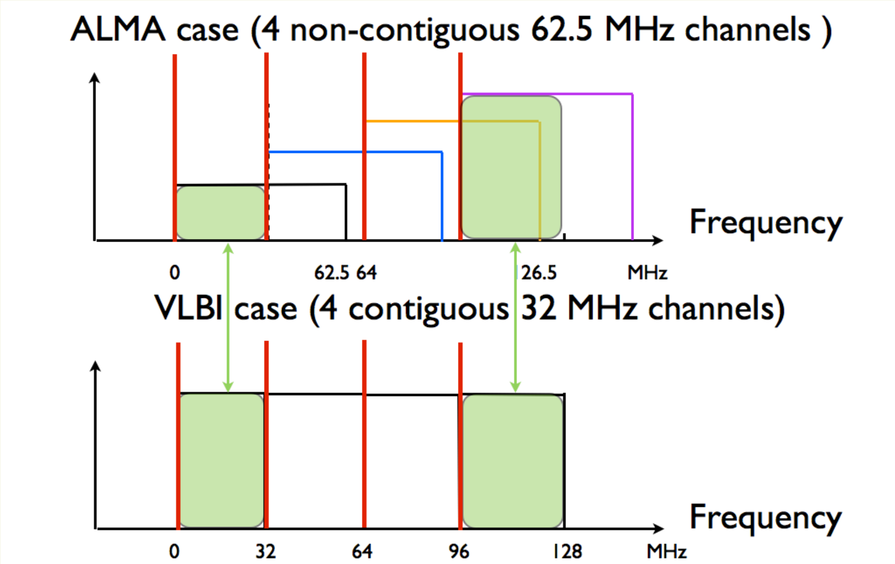
Technical Issues, 4

- Channels/FFT/DFT/zoom-mode/resampling
  - R1DBE: 15 x 32-MHz channels (64 Msps) per IF/pol
  - ALMA: 32 x 62.5-MHz channels (TFB is 125 Msps)
  - R2DBE(now): 1 x 2048-MHz channel (4x1024 Msps)
  - Initial development was aimed for 32/64 v 62.5 MHz
  - R2DBE(future): 2^n-MHz channels? (if needed)
  - R2DBE(future): $2^n 5^m$-MHz channels? (if needed)
  - DBBC3 (in development) will be even more capable
  - ALMA flexibility/1-channel R2DBE allow other options
  - Accommodations to ***difx2mark4/difx2fits***, **fourfit** and possibly AIPS/CASA still remain to be implemented

Zoom Band Automation (Theory)

- From vex file and known configurations of ALMA and other VLBI stations, we derive an optimum configuration of "zoom" band parameters.
- In *v2d*, new parameter in $ZOOM, called *addZoomMatch* with an integer value (1 to 4 presently)
- Automatic selection of zoom band parameters written in *.input* file by ***vex2difx***

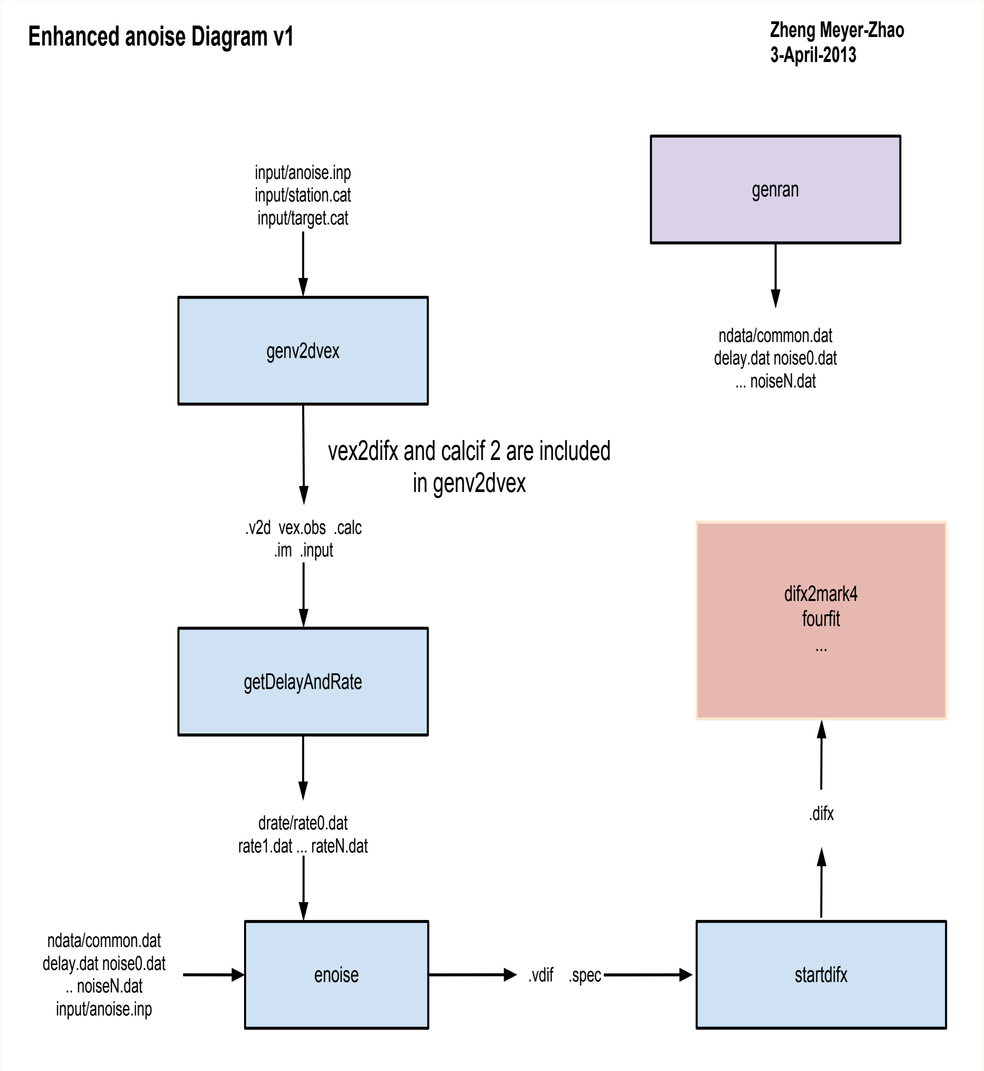




HAYSTACK OBSERVATORY

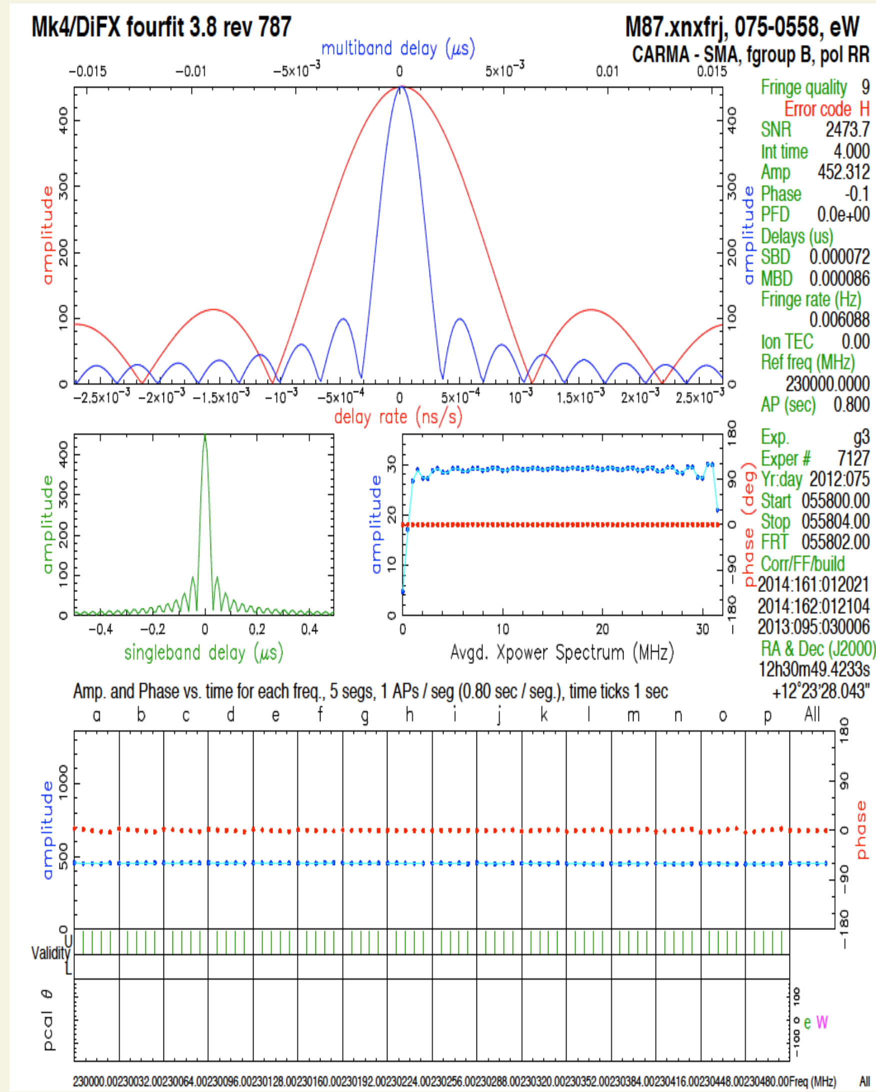
ASIAA VDIF simulator: ***enoise***

- Based on ***anoise*** (prototype, GC), with non-0 baseline option
- Use existing ***calcServer*** for delay model generation
- Input parameter files (vex, v2d,…) of the simulation are created by a shell script
- Can generate complex and realistic configurations with simple control file modifications
  - Stations
  - Sources
  - …



VDIF simulator: ***enoise***

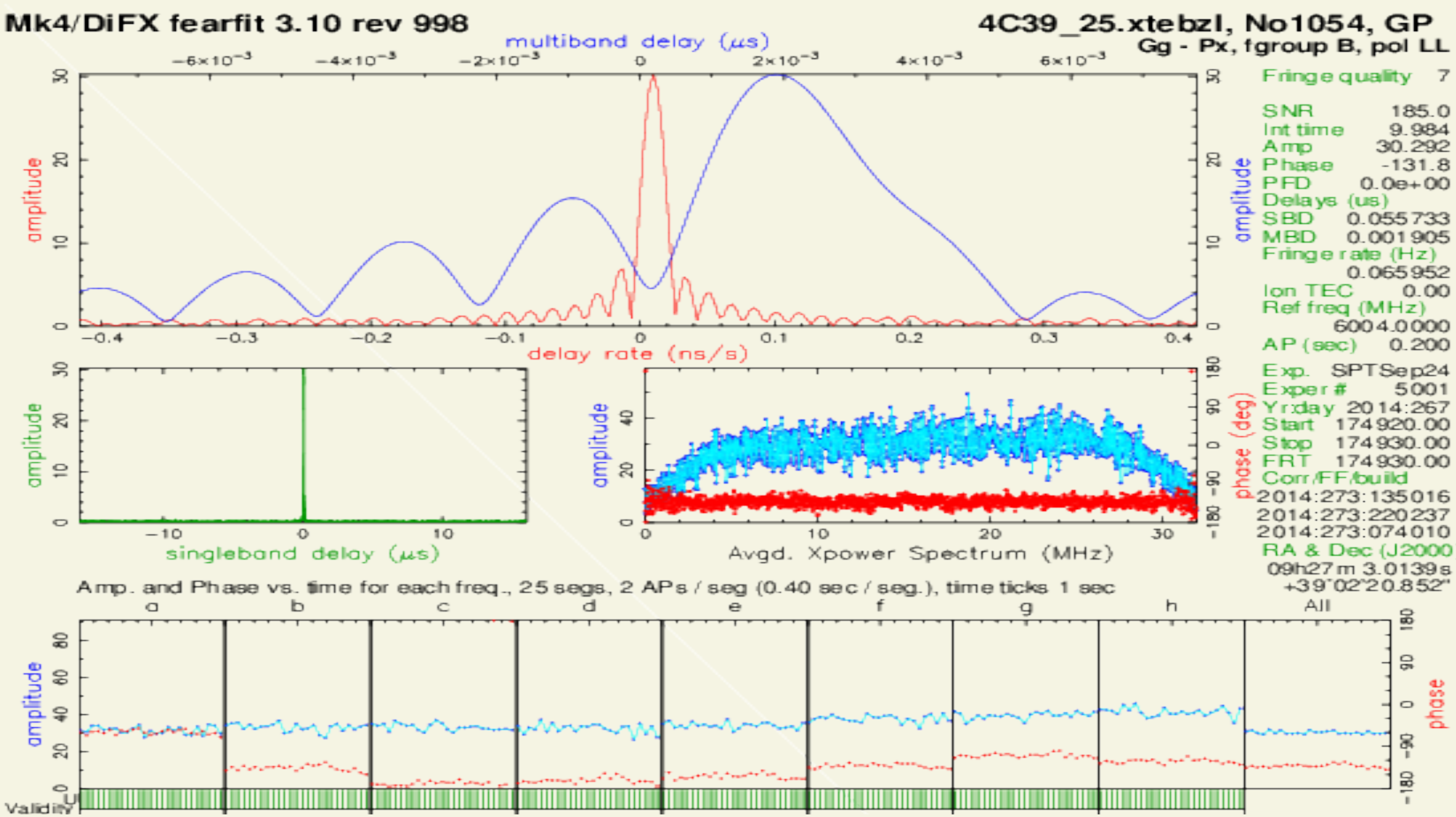
- ***enoise*** can simulate non-0 baseline
  - Any frequency
  - Any baseline
  - Any bandwidth
- Delay model acc. ~100ps
- Used to verify some of the DiFX coding details
- Remaining problems
  - Simulations are very long to compute (~0.1 day per channel per s in APP case)
  - No parallel computing
  - Long duration simulations untested



Technical Issues, 5

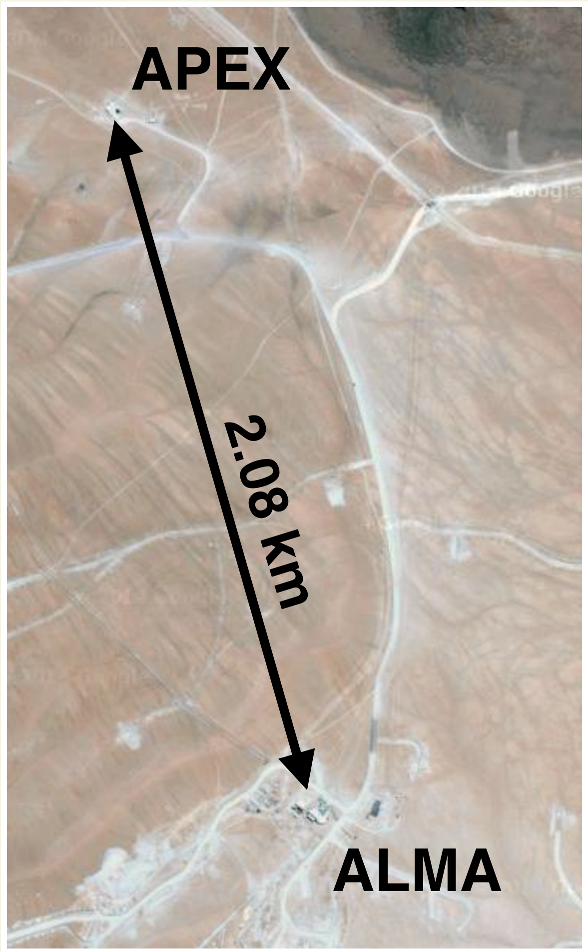
- ALMA sub-band tuning
  - Tunable filter banks tune in 2.0 GHz/2^{16} steps (30.517578125 kHz)
  - Per-sub-band LO offsetting is required in DiFX in order to compensate for the frequency mismatch
  - DiFX support with zoom mode by A. Deller based on work with ASIAA ***enoise*** simulator
  - Small LO offset was tested with SPT equipment (R2DBE) and the Wf/Gg geodetic R1DBE (v1.4) due to a bug in early down-converter implementation

SPT(Wf) R2DBE v GGAO R1DBE



8 Channels of R1DBE zoomed from 2-GHz R2DBE single channel

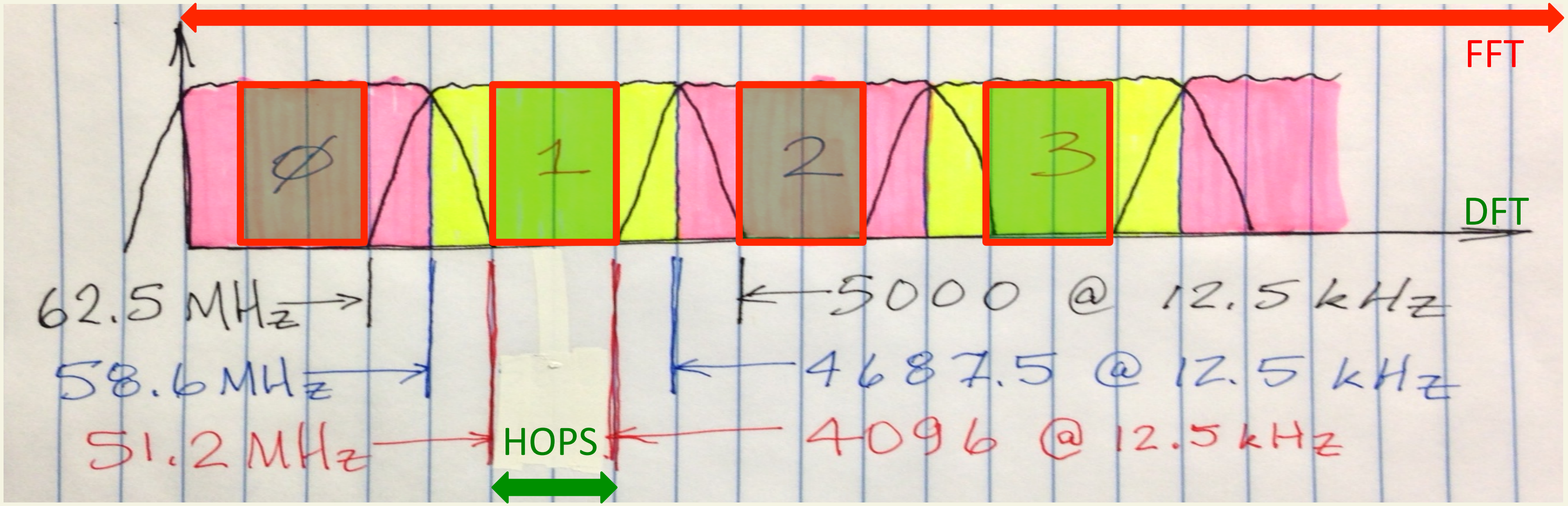
Jan 2015 ALMA-APEX Campaign



- Two nights (lousy weather)
- "un-phased" array (long story)
- 5-min scans on 4 sources
  - (chosen for SPT visibility)
  - 0522-364, 1246-257, 1322-42, 3c279
- 2 2-GHz dual-pol ALMA Bands
- 2 APEX single-pol backends (R2DBE and [modified] DBBC)
- 13 fringes partially analyzed

Reality Intrudes

- ALMA cannot support many spectral windows
- Therefore TFBs are placed for ALMA, not us:
- Channels overlap! Power-of-two for FFT!



DiFX vex $FREQ and .v2d

32 ALMA Channels

```
gbc@gefera:~/APP/trunk/Study/analysis/Commish-Jan
*---------------------- begin $FREQ                ----------------------*
$FREQ;
    def VDIF_8032_2-FREQ;
      * ALMA requested 214989500000.0 Hz
      chan_def = &B : 215928.953125 MHz : L : 62.5 MHz : &Ch01 : &BBC01 : &cp;
      chan_def = &B : 215870.359375 MHz : L : 62.5 MHz : &Ch02 : &BBC01 : &cp;
      chan_def = &B : 215811.765625 MHz : L : 62.5 MHz : &Ch03 : &BBC01 : &cp;
      chan_def = &B : 215753.171875 MHz : L : 62.5 MHz : &Ch04 : &BBC01 : &cp;
      chan_def = &B : 215694.578125 MHz : L : 62.5 MHz : &Ch05 : &BBC01 : &cp;
      chan_def = &B : 215635.984375 MHz : L : 62.5 MHz : &Ch06 : &BBC01 : &cp;
      chan_def = &B : 215577.390625 MHz : L : 62.5 MHz : &Ch07 : &BBC01 : &cp;
      chan_def = &B : 215518.796875 MHz : L : 62.5 MHz : &Ch08 : &BBC01 : &cp;
      chan_def = &B : 215460.203125 MHz : L : 62.5 MHz : &Ch09 : &BBC01 : &cp;
      chan_def = &B : 215401.609375 MHz : L : 62.5 MHz : &Ch10 : &BBC01 : &cp;
      chan_def = &B : 215343.015625 MHz : L : 62.5 MHz : &Ch11 : &BBC01 : &cp;
      chan_def = &B : 215284.421875 MHz : L : 62.5 MHz : &Ch12 : &BBC01 : &cp;
      chan_def = &B : 215225.828125 MHz : L : 62.5 MHz : &Ch13 : &BBC01 : &cp;
      chan_def = &B : 215167.234375 MHz : L : 62.5 MHz : &Ch14 : &BBC01 : &cp;
      chan_def = &B : 215108.640625 MHz : L : 62.5 MHz : &Ch15 : &BBC01 : &cp;
      chan_def = &B : 215050.046875 MHz : L : 62.5 MHz : &Ch16 : &BBC01 : &cp;
      chan_def = &B : 214991.453125 MHz : L : 62.5 MHz : &Ch17 : &BBC01 : &cp;
      chan_def = &B : 214932.859375 MHz : L : 62.5 MHz : &Ch18 : &BBC01 : &cp;
      chan_def = &B : 214874.265625 MHz : L : 62.5 MHz : &Ch19 : &BBC01 : &cp;
      chan_def = &B : 214815.671875 MHz : L : 62.5 MHz : &Ch20 : &BBC01 : &cp;
      chan_def = &B : 214757.078125 MHz : L : 62.5 MHz : &Ch21 : &BBC01 : &cp;
      chan_def = &B : 214698.484375 MHz : L : 62.5 MHz : &Ch22 : &BBC01 : &cp;
      chan_def = &B : 214639.890625 MHz : L : 62.5 MHz : &Ch23 : &BBC01 : &cp;
      chan_def = &B : 214581.296875 MHz : L : 62.5 MHz : &Ch24 : &BBC01 : &cp;
      chan_def = &B : 214522.703125 MHz : L : 62.5 MHz : &Ch25 : &BBC01 : &cp;
      chan_def = &B : 214464.109375 MHz : L : 62.5 MHz : &Ch26 : &BBC01 : &cp;
      chan_def = &B : 214405.515625 MHz : L : 62.5 MHz : &Ch27 : &BBC01 : &cp;
      chan_def = &B : 214346.921875 MHz : L : 62.5 MHz : &Ch28 : &BBC01 : &cp;
      chan_def = &B : 214288.328125 MHz : L : 62.5 MHz : &Ch29 : &BBC01 : &cp;
      chan_def = &B : 214229.734375 MHz : L : 62.5 MHz : &Ch30 : &BBC01 : &cp;
      chan_def = &B : 214171.140625 MHz : L : 62.5 MHz : &Ch31 : &BBC01 : &cp;
      chan_def = &B : 214112.546875 MHz : L : 62.5 MHz : &Ch32 : &BBC01 : &cp;
      sample_rate = 125.0 Ms/sec;
    enddef;
    def VDIF_8224_2-FREQ;
      * ALMA requested 214989500000.0 Hz
      chan_def = &B: 214052.0 MHz: U: 2048 MHz: &CH01 : &BBC01 : &cp;
      sample_rate = 4096.0 Ms/sec;
    enddef;
*----------------------    end $FREQ                ----------------------*
:
```

1 APEX Channel

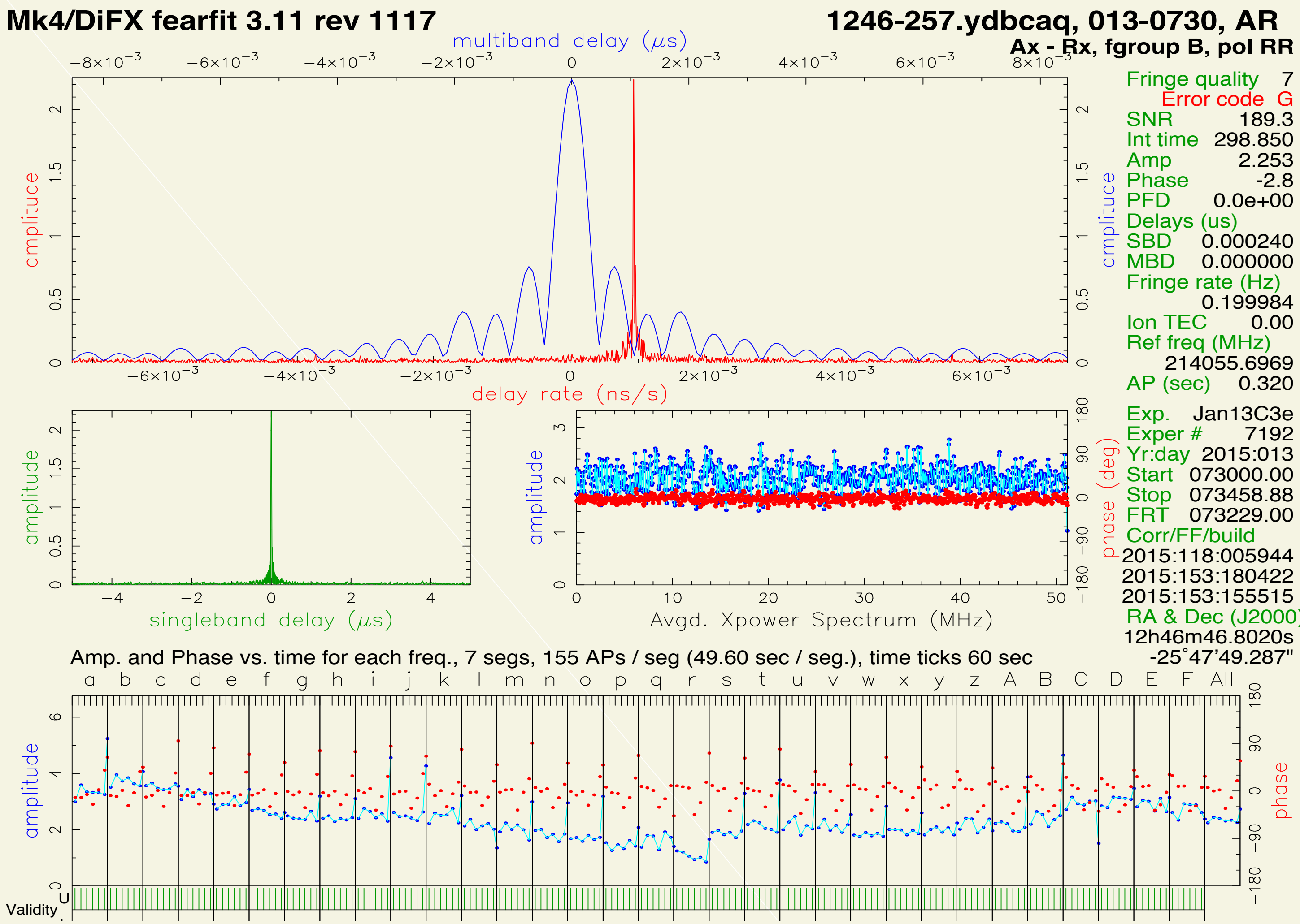
Nasty Numerology

32 Zoom Bands

```
gbc@gefera:~/APP/trunk/Study/analysis/Commish-Jan
        vex = Jan13C3e.vex.obs
        mjdStart = 57035.3125000000
        mjdStop  = 57035.3159722222
        antennas = Ax,By,Rx
        startSeries = 7192
        dataBufferFactor = 16
        visBufferLength = 40
        nDataSegments = 8

        SETUP default
        {
            tInt = 0.32
            subintNS = 32000000
            FFTSpecRes = 0.003125
            specRes = 0.10
            xmacLength = 1
            strideLength = 1
            guardNS = 2000
        }

        ANTENNA Ax
        {
          # dm = rc17
          file = /data-sc02/gbc/alma-data/Q1X/sequences/FLp-/Aa/No1134.vdif
          format = VDIF/8032/2
          phaseCalInt = 0

                addZoomFreq = freq@215872.103125/bw@51.2/noparent@true
                addZoomFreq = freq@215813.509375/bw@51.2/noparent@true
                addZoomFreq = freq@215754.915625/bw@51.2/noparent@true
                addZoomFreq = freq@215696.321875/bw@51.2/noparent@true
                addZoomFreq = freq@215637.728125/bw@51.2/noparent@true
                addZoomFreq = freq@215579.134375/bw@51.2/noparent@true
                addZoomFreq = freq@215520.540625/bw@51.2/noparent@true
                addZoomFreq = freq@215461.946875/bw@51.2/noparent@true
                addZoomFreq = freq@215403.353125/bw@51.2/noparent@true
                addZoomFreq = freq@215344.759375/bw@51.2/noparent@true
                addZoomFreq = freq@215286.165625/bw@51.2/noparent@true
                addZoomFreq = freq@215227.571875/bw@51.2/noparent@true
                addZoomFreq = freq@215168.978125/bw@51.2/noparent@true
                addZoomFreq = freq@215110.384375/bw@51.2/noparent@true
                addZoomFreq = freq@215051.790625/bw@51.2/noparent@true
                addZoomFreq = freq@214993.196875/bw@51.2/noparent@true
                addZoomFreq = freq@214934.603125/bw@51.2/noparent@true
:
```

And similarly for By and Rx

Jan 2015 ALMA-APEX Campaign

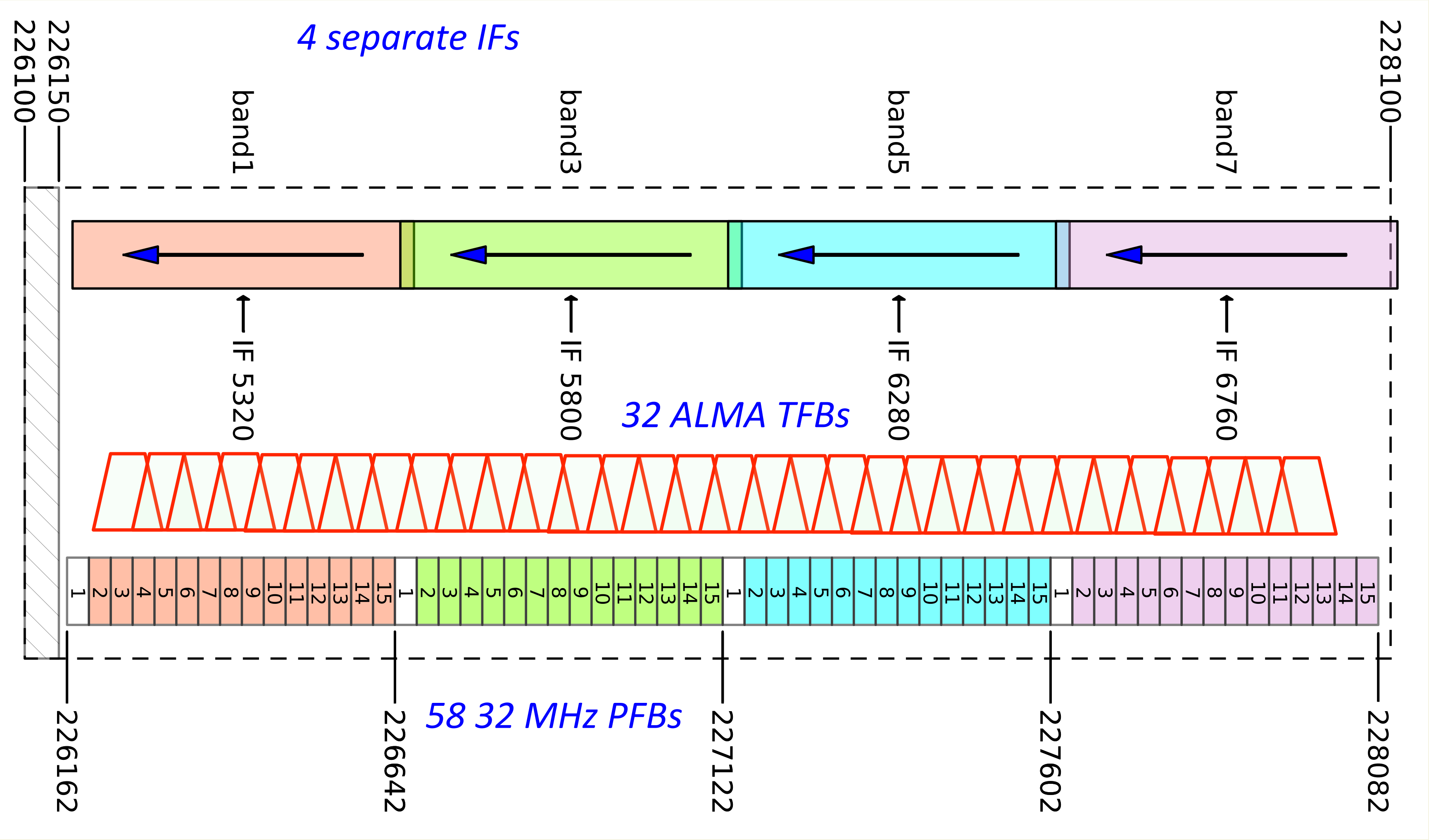


HAYSTACK OBSERVATORY

Mar 2015 ALMA-EHT Campaign

- Blizzard at ALMA
- Recorded 2 (short) scans on 3c273
- Correlation of these are still in progress:
  - Only 9 antennas in phased Array
  - Only SMT, PV and CARMA were observing
  - (CARMA – SMT fringe is very weak)
  - Scan too short for a fringe to SMT?
  - Possibly too weak for PV?
  - CARMA ref data corrupt—requires some work
  - CARMA phased is a challenge…

The Problem w/CARMA-ph-ALMA



Need a Stitching tool

- Channels are incommensurate
- Without rewriting DiFX:
  - Same idea as ***polconvert***
  - Full bandwidth requires ~100 "zoom" bands
  - DiFX output can be stitched back together into new output file
  - Vex, v2d, &c, then re-written as if correlation had been possible
  - Proceed with ***difx2mark4/difx2fits***
- Similar approach needed for ALMA & GMVA

Future Work

- ***enoise*** enhancements as needed for testing
- **Zoom stitching tool** or a significant rewrite….
- May need ***difx2mark4/difx2fits*** enhancements to better accommodate DFT vs FFT and zoom channels (*e.g.* round up # channels to nearest power of 2)
- Modify ***fourfit*** (and other HOPS tools) to use DFT rather than FFT processing (i.e. relax assumptions of 2^n) and modify internal data structures for significantly larger data sets.
- Update script infrastructure for bulk processing