Vienna VLBI and Satellite Software

Post-Correlation Processing (PCP) in VieVS

Jakob Gruber[a]

[a]Technische Universität Wien, Austria

Geodetic VLBI processing

Fundamental data stages in geodetic VLBI

1. Baseband data
2. Visibilities
3. Group delay (systematic errors uncorrected)
4. Group delay (systematic errors corrected)
5. Geodetic parameters (e.g. station velocities, EOP's, ...)

Fundamental processing steps in geodetic VLBI

- 1. → 2. Correlation
- 2. → 3. Fringe-fitting
- 3. → 4. Post-correlation processing
- 4. → 5. Geodetic parameter estimation

Motivation for post-correlation processing

Scheduling	Observing	Correlation	Fringe-Fitting	Post-Correlation Processing	Geodetic Parameter Estimation
TU WIEN VieVS Vienna VLBI and Satellite Software		DiFX	HOPS PIMA	TU WIEN VieVS Vienna VLBI and Satellite Software nuSolve	TU WIEN VieVS Vienna VLBI and Satellite Software

- Additional element in the VLBI processing chain in our group
- More independent data flow in VieVS
- More control over processing algorithms within post-correlation processing

Key questions of this talk

- Which systematic errors need to be corrected in PCP?
- What are the characteristics of these systematic errors?
- Which methods can be used to detect and to correct for them?
- How can we distinguish between those systematic errors?
- How to implement a post-correlation processing toolbox in VieVS?

Systematic errors corrected in PCP

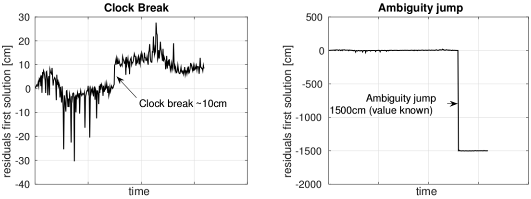
- Group delay obtained from fringe-fitting is affected by systematic errors:
  - Clock jumps (breaks)
  - Ambiguities
  - Ionospheric delay contribution
  - Outliers
- These errors can be in the range of tens to hundreds of ns (100ns=30m)
- Of course there are other systematic affects (e.g. cable delay, geophysical affects)

$$\tau = \tau_{mbd} + n_{amb}\Delta\tau_{amb} + \Delta\tau_{cb} + \Delta\tau_{ion} + (\Delta\tau_{cab} + ...)$$

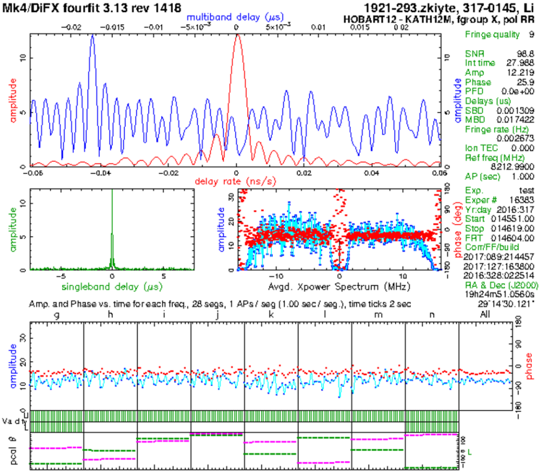
Systematic errors corrected in PCP

$$\tau = \tau_{mbd} + n_{amb}\Delta\tau_{amb} + \Delta\tau_{cb} + \Delta\tau_{ion}$$

- τ_{mbd} ... multi band delay obtained from fringe-fitting [sec]
- $\Delta\tau_{amb}$ ... ambiguity delay size [sec]
- n_{amb} ... integer number of ambiguity
- $\Delta\tau_{cb}$ ... clock jump [sec]
- $\Delta\tau_{ion}$ ... ionospheric delay contribution [sec]
- Θ ... quality flag (due to outliers, bad observations,...)
- τ ... multi band delay free from major systematic affects [sec]



τ_{mbd} multi band delay

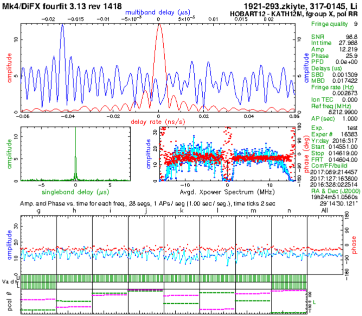


τ_{mbd} multi band delay

- Difference in signal arrival times
- Estimated in the process of fringe-fitting (PIMA, fourfit)
- Stored in particular format:
  - Fourfit Mark4 binary data files (file type-1,-2,-3,4)
  - PIMA ASCII output
- In S/X-system, one multi band delay for X and S-band
- Multi band delay comes with a formal error:

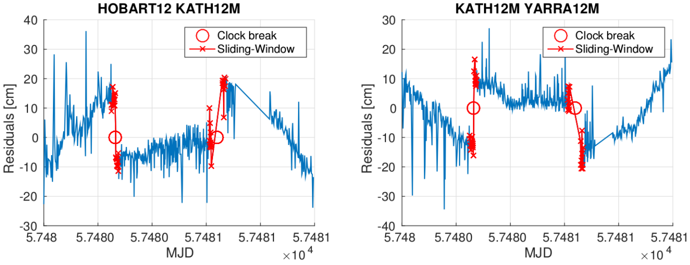
$$\tau_{mbd,X} + \sigma_{mbd,X}$$

$$\tau_{mbd,S} + \sigma_{mbd,S}$$



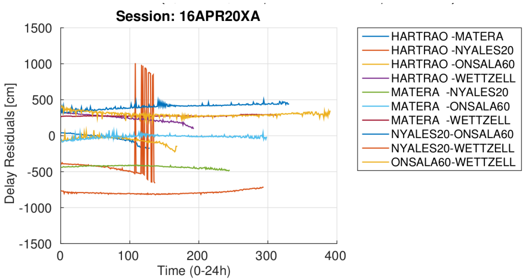
$\Delta\tau_{cb}$ clock jump

- Loss of phase lock on the signal from hydrogen maser
- Jump in time series
- Occurs at station
- Influences all baselines with this particular station (strong condition)



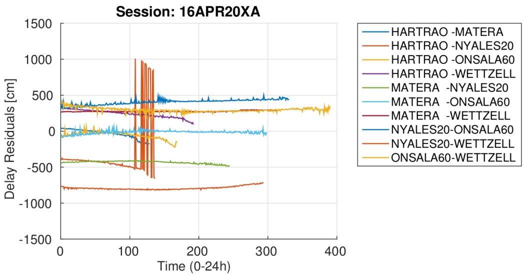
$\Delta\tau_{amb}$ group delay ambiguity

- Multiband delay estimation is based on measurements of phase differences
- Phase difference measurements [rad] are ambiguous by an integer number n_{amb} of the ambiguity delay size $\Delta\tau_{amb}$



$\Delta\tau_{amb}$ group delay ambiguity

- Size of $\Delta\tau_{amb}$ depends
  - on the effective bandwidth for multi band delay $\rightarrow$ frequency setup
  - on the spectral resolution of visibilities for single band delay
- Size of $\Delta\tau_{amb}$ changes with change of frequency setup or loose of channel in correlation/fringe-fitting
- Occurs per baseline
- Size of $\Delta\tau_{amb}$ is exactly known (typically 50/200 ns for X/S-band)



$\Delta\tau_{ion}$ delay due to ionosphere

- Ionosphere is a dispersive medium
- Propagation velocity changes with respect to frequency
- Linear combination of X and S-band multi band delay yields ionosphere free multi band delay
- Higher order terms in the X/S combination are usually dropped in VLBI

$$\tau = \tau_{mbd,X} \frac{f_X^2}{f_X^2 - f_S^2} - \tau_{mbd,S} \frac{f_S^2}{f_X^2 - f_S^2}$$

σ_{mbd} propagation of formal error $\sigma_{mbd,X,S}$

- f_X ... reference frequency for X-band
- f_S ... reference frequency for S-band
- $\tau_{mbd,X,S}$ ... multi band delay from fringe-fitting
- τ ... ionosphere free delay

⊖ quality flag

- To specify if observation will be accepted for geodetic analysis or rejected
- Three cases
  1. Observation is already flagged as bad observation in the fringe-fitting program
  2. Outliers detected in PCP using time series analysis
  3. Data flagging
     - due to bad ionosphere correction (or no S-band available)
     - dropping bad stations, sources, etc.
- Percentages of rejected observations
  - r1854 22%
  - r1855 22%
  - r1856 35%

PCP - realization aspects

- **Theoretically:**

$$\tau = \tau_{mbd} + n_{amb}\Delta\tau_{amb} + \Delta\tau_{cb} + \Delta\tau_{ion}$$

- **Practically:**
  - A database is required to access fringe-fitting output and to store PCP results which can be accessed by geodetic parameter analysis software (e.g. vgosDb, MarkIV)
  - Database conversion from fringe-fitting output db to a database which can be accessed by a PCP-toolbox (e.g. vgosDb)
  - Database update with auxiliary data:
    - cable delay data
    - meteorological data
    - (a priori data for lsm)

PCP with nuSolve

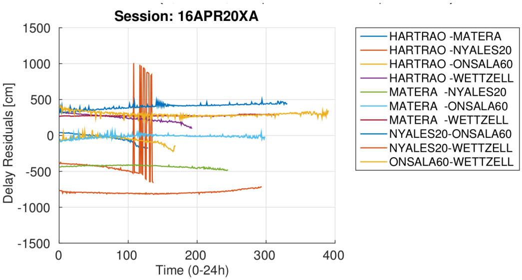
- **The results of the individual PCP processing steps are stored in different levels:**
  - MarkIII: one binary file per processing level (18APR16XA_V001 - V004)
  - vgosDb: wrapper file per processing level in the vgosDb (18AUG09XE_V00i_iIVS_kall.wrp)
- **Practically - realization with nuSolve:**
  - Level 1: database conversion from fourfit output MarkIII to vgosDb (vgosDbmake)
  - Level 2: database update with theoretical values vgosDbCalc
  - Level 3: database update with met data and cable cal from station log files (vgosDbProcLogs)
  - Level 4: PCP corrected (ambiguity, ionosphere) multi band delay stored (nuSolve)
- **Level 4 contains all the data, including a multi band delay free of large systematic affects, which can be used for geodetic parameter estimation**

PCP processing stragety

- **PCP processing stragety to successfully remove systematic influences**
  1. Database conversion to vgosDb
  2. Outliers flagging for single band delay $\tau_{sbd,S,X}$ (no ambiguities, outliers can be detected more easily)
  3. Clock break correction, ambiguity correction of multi band delay $\tau_{mbd,S,X}$
  4. Outliers flagging for multi band delay $\tau_{mbd,S,X} \rightarrow 3$
  5. Ionospheric correction
  6. Store PCP results in vgosDb

Methods for ambiguity correction

- Visual detection and correction
- Closure condition (at least a network of six station is required to detect affected baselines)
- Iterative processing (Kareinen et al. (2016), PCP talk from last year)



PCP in VieVS

- Database conversion tool (Matlab tool vgosDbMake.m to convert fringe-fitting output to vgosDb)
- Access database (VieVS can access vgosDb)
- Possibility to access certain observations, e.g. $\tau_{X,mbd}$ (select certain parameters in vgosdb_input_settings.txt)
- Estimation of delay residuals to detect systematic influences (can be carried out by VieVS first solution)
- Outliers flagging (tool to flag outliers exists, but is not designed for iterative approaches because very time consuming and it does not allow more sophisticated data flagging methods)
- Clock break detection and correction (VieVS can consider clock breaks)
- Ambiguity detection and correction (open task)
- Ionospheric correction (open task)

PCP in VieVS

Database conversion and update

- Tool to convert the fringe-fitting output to vgosDb: createvgosdb.m
- **Realization:**
  - Library with matlab and bash scripts to read fringe-fitting output and station log files and create a vgosDb
  - Makes heavy use of NetCDF library, supports NetCDF version 4.3.3.1
- **Input:**
  - fourfit output files (type-1,2,3,4)
  - PIMA ASCII output
  - Station log files (configuration file for exception handling, e.g. sign of cable delay)
- **Verification:** Comparison with nuSolve tool vgosDbMake of correlated AUA sessions
- **Application:** Has been used for the European intensive session v012 (eint01)

PCP in VieVS
Data flagging

- Data flagging plays an essential role in PCP
- Several iterative lsm adjustments are necessary in PCP
- Move of data flagging routines from VIE_INIT to VIE_LSM
- With this we avoid running the time consuming part VIE_MOD
- What happens when dropping stations or sources? Affects VIE_MOD?
- Use of more sophisticated data flagging concepts within the vgosDb instead of the current outlier file with MJD time ranges

Key questions of this talk

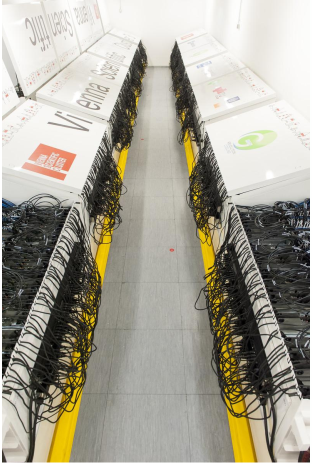
- Which systematic errors need to be corrected in PCP?
  - Clock break, ambiguities, ionosphere and also outliers
- What are the characteristics of these systematic errors?
  - Size of tens to hundreds of nanosec
  - Occur per station or only per baseline
  - Jumps in time series
  - Systematic offset of whole time series
- Which methods can be used to detect and to correct for them?
  - Visual detection
  - Use of change detection algorithms, help of closure conditions
- How can we distinguish between those systematic errors?
  - Use a of particular processing strategy
- How to implement a post-correlation processing toolbox in VieVS?
  - Move data flagging to VIE_LSM
  - Use of first solution to systematic influences

History of Correlation in Vienna

- 2014
  - Installation of the VSC-3
- 2016
  - Jun: Installation of DiFX and HOPS on the VSC-3 (J.McCallum)
  - Jun: Correlation of first session AUG032
  - Sep: Correlation of first satellite observation APOD2
  - Nov: Correlation of McWz
- 2017
  - May: Correlation of ds317
  - Sep: Correlation of first official IVS session AUA025 (J.Gruber, J.McCallum)
  - Dec: Correlation of AUA026
- 2018
  - April: Correlation of European Intensive Session
  - Correlation of CRDS94, verified by WACO
  - Correlation of AUA028, AUA032, ...
  - Correlation of SBL500

Hardware Capabilites of the VSC-3

- **Hard facts:**
  - Installed by ClusterVision
  - Consists of 2020 nodes (each equipped with 2 processors), 2020 nodes equals to 32320 cores
  - Intel Xeon E5-2650v2 processors from the Ivy Bridge-EP family with 8x2.60 GHz and 20MB SmartCache
  - Intel QDR-80 dual-link high-speed InfiniBand fabric
  - BeeGFS parallel Filesystem
- **Suitable for VLBI correlation?**
  - BeeGFS parallel Filesystem developed for intensive i/o workloads makes the VSC-3 to an appropriate computing environment for the VLBI correlation



VLBI Correlation at the VSC-3

- **Current resources at our institute:**
  - 230 TB private storage media
  - Cores available upon request, usually up to 160 cores
- **Characteristics:**
  - The VSC is maintained by IT experts
  - Support by VSC-3 IT experts when installing new software
  - Processing power (number of nodes) can be adjusted to the current correlation work load
  - Linked to a very high data rate connection (GÉANT)

VIENNA SCIENTIFIC CLUSTER

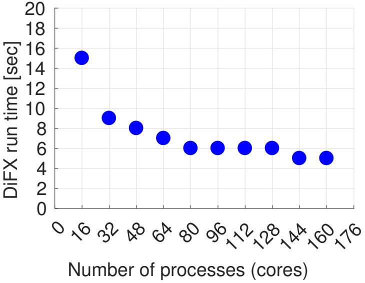
Software Capabilities of the VSC-3

Data Transfer	Correlation	Fringe-Fitting	Post-Correlation Processing
jive5ab	DiFX	HOPS	nuSolve

- **jive5ab** for data transmission
- Distributed FX (**DiFX**) software correlator
- Haystack Observatory Postprocessing System (**HOPS**) with the program fourfit for fringe-fitting
- **nuSolve** is used for post-correlation processing and for geodetic parameter estimation to verify the correlation results
- **SLURM** for cluster management and job scheduling

Performance Tests per scan

- To evaluate the most efficient VSC-3 node configuration per scan
- Methodology: Rerun of the same scan with the same DiFX processing with changing number of cores:
  - Accumulation period 0.512 sec, spectral resolution 0.25 MHz
  - 1 to 10 nodes with 16 processes (cores)
  - session v012, scan 060-1500b, stations SA,WN
- Significant improvement until 80 cores



Performance Tests of a session

- Usage of parallel scan processing strategy to make efficient use of more cores.
- Can be realized with the **SLURM job array**
- For example: 320 cores are available. Split up the pool of cores into several DiFX jobs in parallel. Each scan takes 6 sec.

Serial scan processing

Total processing time 24 sec

- scan 1 — 320 cores
- scan 2 — 320 cores
- scan 3 — 320 cores
- scan 4 — 320 cores

Parallel scan processing

Total processing time 6 sec

- 80 cores — scan 1
- 80 cores — scan 2
- 80 cores — scan 3
- 80 cores — scan 4

Data Transfer from and to the VSC-3

- The VSC-3 can be accessed via five IP addresses
- On each of the IP addresses runs a jive5ab process
- There is one UDP/TCP port per login node
- Pushing data
- Pulling data
- White-listing is a two step process which takes some time

	whitelisted	**tested**
Bonn	~	x
Hartrao	✓	✓
INAF	✓	✓
Ny-Alesund	✓	✓
Onsala	✓	x
Japan	✓	✓
UTAS	✓	✓
Warkworth	✓	✓
Wettzell	✓	✓

Sessions correlated AUA sessions

- **Scientific background:** VLBI SOuthern Astrometry Project (SOAP) program
- **Stations:** AuScope Array with HOBART26, KATH12M, YARRA12M, HARTRAO, HART15M, WARK12M, WARK30M, DSS34
- **Antenna setup:** 1 Gbps recording rate and S/X-band
- **Results:**

Session Name	QCODE	% of Total scans
AUA025	5-9	54%
AUA026	5-9	58%
AUA028	5-9	71%
AUA032	5-9	70%

- **Characteristics:** Absence of FLUX information of sources, G-codes on local baselines HARTRAO-HART15M and WARK12M-WARK30M, X-band only for WARK30M

Sessions correlated

European intensive sessions

- **Scientific background:** ESA Project "Independent Generation of Earth Orientation Parameters"
- **Stations:** WETTZ13N, RAEGSMAR, WETTZELL
- **Antenna setup:** 1 Gbps (256 Mbps) recording rate and S/X-band
- **Results of v012:**
  - Detection of a few S-band fringes due to the lower recording rate observed than scheduled
  - Determination of a high number of X-band multi-band delay observations
  - SBD residuals scatter of ~20 ns due to a priori station coordinates of REAGSMAR
- **Status of EINT02,03,04,...,07:**
  - Detection in the S-band, no detections in the X-band
  - Further investigations are following

VieVS Post-Correlation Processing

- **Motivation:** To close the gap between the correlation/fringe-fitting output and VieVS
- **Components:**
  - Database conversion and update
  - Ambiguity correction
  - Ionospheric delay correction
  - Outliers flagging

VieVS Post-Correlation Processing status

- **Motivation:** To close the gap between the correlation/fringe-fitting output and VieVS
- **Components:**
  - Database conversion and update ✓
  - Ambiguity correction In work using closure conditions
  - Ionospheric delay correction On to-do list
  - Outliers flagging In work conversion of VieVS outlier file to vgosDb

VieVS Post-Correlation Processing Database conversion and update

- Tool to convert the fringe-fitting output to vgosDb: createvgosdb.m
- **Realization:**
  - Library with matlab and bash scripts to read fringe-fitting output and station log files and create a vgosDb
  - Makes heavy use of NetCDF library, supports NetCDF version 4.3.3.1
- **Input:**
  - fourfit output files (type-1,2,3,4)
  - PIMA ASCII output
  - Station log files (configuration file for exception handling, e.g. sign of cable delay)
- **Verification:** Comparison with nuSolve tool vgosDbMake of correlated AUA sessions
- **Application:** Has been used for the European intensive session v012

Research Pipeline

Scheduling	Observing	Correlation	Fringe-Fitting	Post-Correlation Processing	Geodetic Parameter Estimation
TU WIEN VieVS Vienna VLBI and Satellite Software		DiFX	HOPS PIMA*	TU WIEN VieVS Vienna VLBI and Satellite Software	TU WIEN VieVS Vienna VLBI and Satellite Software

- Independent data flow in VieVS
- Analysis of impact of correlation/fringe-fitting models and configurations on geodetic parameter estimation
- *PIMA fringe-fitting is carried out by H.Krásná

Simulation of VLBI Baseband Data

- Software to generate simulated digitized VLBI baseband data
- **Input parameters:** Source strength, antenna sensitivity, ...
- **Output:** Digitized bit stream for each channel per station separately
- **Purpose:**
  - Knowledge gain, better understanding and insight of important parameters of the baseband data of a station
  - Can be used as input data for a correlation/fringe-fitting prototype
  - (Simulation) studies
- **Realization details:**
  - Written in Matlab
  - Bit stream is already split up in accumulation periods in the baseband data generation
  - FFT interpolation to make it possible to add delays with higher precision than recording rate

Simulation of VLBI Baseband Data

- Input parameters can be defined per channel and per accumulation block:
  - Sampling rate/bandwidth
  - Scan length
  - Integration time/spectral resolution
  - Source strength, antenna sensitivity
  - Phase cal repetition rate, pulse amplitude
  - Antenna velocity, source velocity
  - Sky frequency
  - Filter design
  - Cable delay
  - Channel phase delay
  - VLBI geometrical delay τ
  - Source structure, beam width
  - Bit depth
  - (Test signal for verification)

Simulation of VLBI Baseband Data

Input parameters

- Input parameters can be defined per channel and per accumulation block:
  - Sampling rate/bandwidth ✓
  - Scan length ✓
  - Integration time/spectral resolution ✓
  - Source strength, antenna sensitivity ✓
  - Phase cal repetition rate ✓ , pulse amplitude common value?
  - Antenna velocity, source velocity ✓
  - Center frequency ✓
  - Filter design type: ideal ✓ realistic amplitude shape, Chebyshev?
  - Cable delay ✓
  - Channel phase delay ✓
  - VLBI geometrical delay τ ✓
  - Source structure, beam width Not implemented yet!
  - Quantization bit depth Details about quantization? Cut off amplitude? Equidistant bit ranges?
  - (Test signal for verification) ✓

Outlook

- Further investigations of DiFX performance improvement on the VSC
- VSC-4 will be set up early 2019
  - We will gain access to dedicated storage 1 PB and 250 cores
- Ongoing correlation of AUA experiments
- Further development of post-correlation toolbox in VieVS
- Development of a correlation/fringe-fitting prototype in Matlab
- Spread the correlation knowledge to all members of the TU Wien VLBI group:
  - Johannes Böhm
  - Sigrid Böhm
  - Andreas Hellerschmied
  - David Mayer
  - Matthias Schartner

contact: jakob.franz.gruber@geo.tuwien.ac.at