VLBI Correlation Activities at TU Wien

Jakob Gruber [1], Johannes Böhm [1], David Mayer [1], Jamie McCallum [2]

Abstract Geodetic VLBI correlation poses a new challenge in the current VLBI activities at the research area Higher Geodesy at Technische Universität Wien (TU Wien). We are using the Distributed FX (DiFX) software correlator and the Haystack Observatory Postprocessing System (HOPS) on the Vienna Scientific Cluster 3 (VSC-3), which is a supercomputer located at TU Wien. We provide more technical details about the VSC-3 and information about activities related to correlation at TU Wien. Furthermore, we present tools to directly access the correlation and fringe-fitting output database with the Vienna VLBI and Satellite Software (VieVS) using the vgosDb format and we discuss post-correlation processing aspects in VieVS based on currently correlated experiments.

Keywords VLBI correlation, Vienna Scientific Cluster

1 Introduction

VLBI correlation refers to the process of determining the difference in signal arrival times at two stations by comparing the recorded bit streams. These bit streams are characterized by very high sampling rates, ranging from several hundred Mbps to a few Gbps. Thus, for the realization of VLBI correlation, a high performance computing environment is convenient to carry out the correlation task efficiently and a large disk space is required to store all the recorded bit streams of various VLBI stations. Additionally, a stable link with high data rates is needed to transfer the data to the correlator. At TU Wien, we are capable of scheduling VLBI observations and estimating geodetic parameters with the Vienna VLBI and Satellite Software VieVS [1]. With the installation of a VLBI correlation infrastructure, we establish an additional element in the VLBI processing chain (see Table 1) allowing us to contribute with a wider field of capabilities to the VLBI community. This will become even more important with the increasing amount of data due to the VGOS observing scenario.

We installed the Distributed FX-style Correlator (DiFX [3]) on the Vienna Scientific Cluster 3 (VSC-3), a collaboration of several Austrian universities that provides supercomputer resources and corresponding services to their users. Technical details and inital performance tests of the VSC-3 are described in [2]. Here, we focus on data transfer capabilities, further performance studies of the VSC-3, the VLBI data processing pipeline at TU Wien, and current correlation activities.

Table 1 Core tasks in the VLBI processing chain with current and planned realizations.

VLBI task	current status	plan
Scheduling	VieVS	VieVS
Observation	–	–
Correlation	DiFX	DiFX
Fringe-fitting	HOPS/PIMA	HOPS/PIMA
Post-correlation	νSolve	VieVS
Analysis	VieVS	VieVS

1. Technische Universität Wien
2. University of Tasmania

2 Data Transfer Capabilities of the VSC-3

The VSC-3 is located in Vienna and linked to the Gigabit European Academic Network (GEANT). GEANT is a pan-European data and communication network that spans connections to 38 countries in Europe and other regions in the world with data transfer rates up to several backed-up 10-Gbps links. The VSC-3 can be accessed by five login nodes sharing the 10-Gbps link, which means that each of the login nodes of the VSC-3 can be accessed with 2 Gbps on average. For the data transfer of VLBI baseband data we make use of the jive5ab software package developed by H. Verkouter.

To avoid several data transfers through the same login node, we split up the data transfer into streams (usually one per station) and transfer the data through several login nodes in parallel to make efficient use of the VSC-3 network infrastructure. One of our main tasks of the VSC-3 is the correlation of data from the AUSTRAL network. For the data transfer between Hobart and Vienna, the pull method is used to retrieve the data and it is split up into several data streams—one per login node. With this concept, any bottlenecks of transmission via a single login node in Vienna are avoided and a total download rate of 1.8 Gbps from Hobart can be achieved effectively.

3 Correlation Processing Performance of the VSC-3

High performance clusters like the VSC-3 offer a large amount of cores which can be used on a shared basis with other users for processing intensive tasks, e.g., VLBI correlation. To some extent, the user can decide how many cores should be used for the processing task. As shown in [2], the correlation software DiFX scales up to approximately six nodes at the VSC-3 for an AUSTRAL session; therefore, we usually request six nodes for the correlation of those sessions because the use of more nodes does not decrease the processing time. The DiFX software architecture is well designed to efficiently process the baseband data of all stations that contribute to only one single scan.

However, to make efficient use of more cores we test a parallel scan processing strategy. While in a serial scan processing strategy all the processing power is assigned to one single scan, a parallel scan processing strategy spreads all the processing power efficiently to several scans.

3.1 Methodology and Realization

In contrast to the MPI parallelization within one single scan, parallelization of several scans can be realized with the concept of job arrays provided by the Slurm workload manager. In the method described here, each job of the Slurm array refers to the execution of a DiFX job for a single VLBI scan. The Slurm job array allows the execution of several jobs in parallel making it possible to correlate several VLBI scans in parallel and make efficient use of more cores. In this performance test we change the number of scans that are correlated simultaneously to evaluate possible impacts on the correlation processing time. The relation of the number of scans processed in parallel and the correlation processing time is estimated. Furthermore, this test reveals if the network data traffic and disk work load reach a critical point where latency becomes an issue and the VSC-3 network and disk infrastructure cannot be used anymore within a certain test setup.

3.2 Test Setup

We use real observed data from two single baseline experiments: a four-hour session with the local baseline in HartRAO called SBL500 and a one-hour intensive-style session. Since SBL500 was observed in 2-Gbps recording mode, it is a very handy experiment to evaluate the data load on the system due to the very high recording rate. Only scans with 30-sec duration are used to ensure consistency between the scans which are processed in parallel. This means that 66 scans were used for correlation. In the DiFX correlation setup, 1-sec integration times are selected along with 128 channels for the spectral resolution. The second experiment used in this performance test is a one-hour VLBI session with the new VGOS antennas in Wettzell and Santa Maria with 73 scans (session v012). This session is dedicated to an ESA project for the independent generation of Earth Orientation Parameters (EOP). Since this kind of session has an even greater value when analyzed in near real-time, testing the correlation time performance for this session has important practical reasons. It is recorded with 256 Mbps and consists of mainly 20-sec scans. In the DiFX correlation setup, a 1-sec integration time is selected and 64 channels are used for the spectral resolution.

3.3 Results

In this performance test the impact of the additional workload due to simultaneously loaded data of parallel scan processing is evaluated. A clear correlation between number of scans processed in parallel and the processing time can be found (see Figure 1).

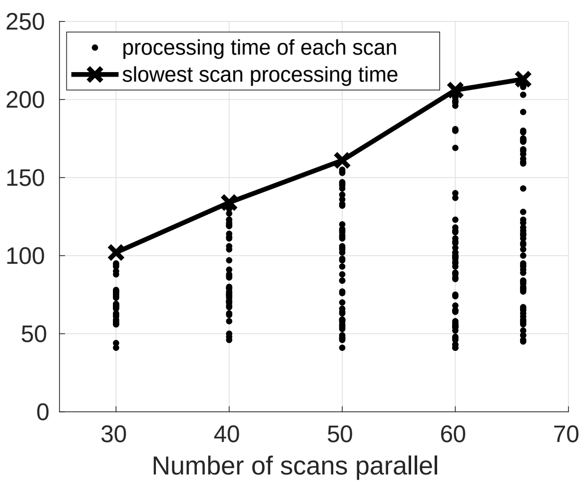

Fig. 1 Processing time. DiFX processing time on the VSC-3 with respect to number of scans which are processed in parallel using the SLURM job array.

This might be mainly due to the additional data traffic in the VSC-3 network and due to the increased reading processes on the BeeGFS file system of the VSC-3. While this test using the SBL500 session basically shows that the VSC-3 is capable of processing a high number of scans in parallel (in this case up to 66 scans), it also reveals that the total processing time increases as well. This means that the amount of data which is processed per unit time does not increase effectively. The effective throughput is characterized by the processing rate, which shows how much data is processed in one second. The processing rate for the SBL500 session is depicted in Figure 2 and shows that there is no improvement in data rates when using an increased number of scans processed in parallel.

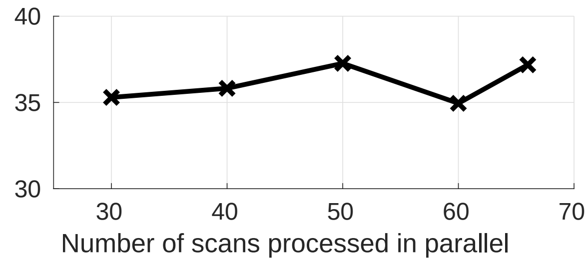

Fig. 2 Processing rate. Relation between processing rate and number of scans processed in parallel.

The data rate performance is flat with a value of around 36 Gbps between 30 and 66 scans in parallel. This means there is no increase in efficiency when going from 30 to 66 scans in parallel as almost the same amount of data is processed in the same time. However, it still needs to be evaluated at which point the data rate performance does not improve any more. There might be a point for a certain number of scans processed in parallel below 30 at which the VSC-3 stops scaling. Below this number the data rate might efficiently increase with an increasing number of scans until it reaches the 36-Gbps level. In any case, with a total rate of 36 Gbps the VSC-3 is also capable to process large VLBI networks with high recording rates in real-time. Furthermore, this test shows that the VSC-3 is capable of processing in total 990 GB (66 scans x two stations x 2 Gbps x 30-sec scans) at once without completely overloading the hard drive and the network.

Of great interest is the parallel scan strategy for intensive sessions because we want to keep the latency of the delivery of correlation results as short as possible. For this purpose we apply the parallel scan processing strategy for the v012 session and run all scans of this session in parallel. Usually, the processing time for a single scan of this certain recording setup with one node takes 16 sec for processing. Due to the small total amount of loaded data in comparison to the SBL500 session with 2 Gbps, the parallel scan processing strategy is very effective for one-hour sessions such as the v012 European intensive session. Using the parallel scan processing strategy a total processing time for the whole session of 22 sec can be achieved. With this value we completely remove the bottleneck for real-time processing of VLBI intensive sessions from the correlation task, and other tasks are crucial for a fast provision of intensive session results like data transfer speed.

4 VLBI Data Processing Pipeline at TU Wien

Once the incoming radiation is received, digitized by the VLBI stations, and transmitted to the correlator such as the VSC-3, several processing steps need to be carried out to finally obtain geodetic parameters, e.g., station coordinates, Earth orientation parameters, and atmospheric parameters. The processing steps can be split up into four core tasks. First the digital baseband data needs to be correlated to generate so-called visibilities. Thereafter, a multiband delay is estimated out of several single-band visibilities. This process is called fringe-fitting. In the legacy S/X system a multiband delay is estimated per band (X- and S-band). However, these observations are affected by large systematic errors (several tens to hundreds of nanoseconds) including clock jumps, unresolved ambiguities, and delays due to the dispersive medium (the ionosphere). The correction of those systematic errors is carried out within the task of post-correlation processing. In this step essential database updates such as the addition of cable delay information and meteorological parameters and data flagging are carried out. Post-correlation processing yields the fundamental observations which are free of large systematic influences and then used for geodetic parameter estimation. Several individual software packages are required to carry out all the processing steps from baseband data to the final geodetic parameters. At TU Wien, we implemented DiFX for correlation, HOPS and PIMA for fringe-fitting, and νSolve for post-correlation processing in our working environment in addition to our Vienna VLBI and Satellite Software (VieVS). Currently, we use νSolve for post-correlation processing and geodetic parameter estimation to verify the correlation and fringe-fitting results because it provides valuable insights of the correlation/fringe-fitting results.

Additionally, we are working on a post-correlation processing toolbox to directly access the fringe-fitting output with VieVS. We have developed a tool to convert the fringe-fitting output to vgosDb. This tool is capable of reading fourfit binary output files (type-1, 2, 3, 4) as well as PIMA ascii output, and it produces a vgosDb file which also takes the cable delay and meteorological information from station field log files into account. It is realized mainly with Matlab and makes heavy use of the NetCDF library. A verification was carried out with the νSolve tool vgosDbMake of correlated AUA sessions and it was applied to the analysis of the European intensive sessions. Other components, which will be implemented in this post-correlation toolbox, are ambiguity correction algorithms, ionospheric delay correction and data flagging tools. With such a toolbox we can realize a more independent data flow in VieVS and we gain experience with algorithms that are necessary for the post-correlation processing tasks. Furthermore, we can better evaluate the impact of correlation/fringe-fitting models and configuration setups on geodetic parameter estimation.

5 Correlation Activities at TU Wien

Several different types of VLBI sessions were correlated, mostly for scientific geodetic and astrometric purposes, such as official IVS sessions dedicated to the VLBI SOuthern Astrometry Project (SOAP) using the AuScope array plus the local baselines at Hartebeesthoek and Warkworth. For the first time both New Zealand antennas participated simultaneously. Those sessions are correlated on a monthly basis and the results are published via the IVS. Besides the AUA sessions from the SOAP program, satellite observations with VLBI telescopes were correlated as well as other specific VLBI sessions such as the short baseline session (SBL500) using the local baseline at Hartebeesthoek. Furthermore, we carry out the correlation of one-hour Intensive sessions with the northern VGOS antenna at Wettzell and Santa Maria. These sessions represent a reasonable test case to work on an highly automated processing pipeline to provide geodetic products in near real-time. This involves methods and algorithms to process the data with minimized manual interactions on the one hand, and on the other hand it uses the total potential of transfer speed and correlation capabilities as described in Sections 2 and 3. We also want to note here that a successful comparison against the Washington correlator was carried out.

6 Conclusion and Outlook

We describe various fields of work with respect to VLBI correlation activities at TU Wien. In particular, we show data transfer capabilities of the 10-Gbps links of the VSC-3 into the GEANT network, with a tested download rate of 1.8 Gbps from a flexbuff at the University of Tasmania. A method for parallel scan processing with DiFX is described to efficiently use the processing power of a supercomputer like the VSC-3. With such a method it is possible to achieve a total correlation processing throughput of up to 36 Gbps.

To get a more independent data flow we are working on a post-correlation processing toolbox in VieVS. At this point we can carry out a data base conversion from the fringe-fitting output of fourfit and PIMA to vgosDb. Furthermore, we are correlating official IVS sessions on a regular basis from the AUSTRAL network. The correlation results of the VSC-3 were successfully compared against the correlation results of the Washington correlator, which shows no significant differences. At the moment we are in the progress of becoming an official IVS Correlation Center and we will carry out further correlation of IVS sessions on a regular basis in the future. As a university we are also interested in research topics within the correlation, fringe-fitting, and post-correlation tasks. We are highly interested in data processing of VGOS data and we plan to develop a generator for simulated digitized VLBI baseband data in the VDIF format.

In early 2019, we will gain access to dedicated storage of 1 PB and 250 private cores on a new realization within the Vienna Scientific Cluster family (VSC-4). We will carry out performance tests and efficiency analyses on the VSC-4 as well and we will work on a refined processing pipeline. Finally, we are working on spreading the correlation knowledge to all members of the TU Wien VLBI group.

Acknowledgements

The correlation has been carried out on the Vienna Scientific Cluster (VSC-3). We acknowledge the Austrian Science Fund (FWF) for supporting our work in projects SORTS (I 2204) and VGOS Squared (P 31625).

References

1. J. Böhm, S. Böhm, J. Boisits, A. Girdiuk, J. Gruber, A. Hellerschmied, H. Krásná, D. Landskron, M. Madzak, D. Mayer, J. McCallum, L. McCallum, M. Schartner, K. Teke. Vienna VLBI and Satellite Software (VieVS) for Geodesy and Astrometry. *Publications of the Astronomical Society of the Pacific*, 130(986), 044503, 16, 2018.
2. J. Gruber, J. Böhm, and J. McCallum. Geodetic VLBI Correlation at the Vienna Scientific Cluster. In R. Haas and G. Elgered, editors, *Proceedings of the 23rd European VLBI Group for Geodesy and Astrometry (EVGA) Working Meeting*, Chalmers University of Technology, 2017, ISBN: 978-91-88041-10-4, pages 140–144.
3. A.T Deller, S.J. Tingay, M. Bailes, C. West. DiFX: A Software Correlator for Very Long Baseline Interferometry Using Multiprocessor Computing Environments. *Publications of the Astronomical Society of the Pacific*, 119, 318336, 2007.