Bonn correlator report to the TOG 2010

Introduction: DiFX Software Correlator

In Bonn VLBI observations are now processed using the **DiFX software correlator**. DiFX (Distributed FX correlator) was developed at Swinburne University in Melbourne, Australia by Adam Deller (and collaborators), and adapted to the VLBA operational environment by Walter Brisken and NRAO staff.
The present installation at the MPIfR in Bonn realized by Helge Rottmann followed a somewhat different path, with the intention to switch to a system similar to NRAO's as soon as the software limitations of the Mark 5 systems (only 32-bit API available) will allow this. DiFX correlators have been installed at Swinburne and ATNF (Australia), NRAO (USA), Bonn (MPIfR) for production correlation and a few other places for more experimental projects.
Software correlation has become feasible in recent years, and is especially well suited to applications like VLBI with bandwidth-limited data-transmission systems and non-real-time processing. Among its several **advantageous aspects** are:

- flexible allocation of processing resources to support correlation of varying numbers of stations, frequency and time resolution, and various special processing modes, with no fundamental fixed limits other than the finite performance of the processing cluster
- optimization of resource usage to minimize processing time
- integration of control and processing functions
- continuously scalable, incremental upgrade paths
- relatively straightforward implementation of special modes and tests

DiFX processes **2-bit samples** with substantially greater efficiency than 1-bit samples over double the bandwidth, basically because only half as many samples must be correlated. Since these two cases have nearly equivalent sensitivity, it is recommended to record in 2-bit mode.
Operation of DiFX is governed primarily by an observation description in **VEX format**. This format is used for both station and correlator control functions in a number of VLBI arrays, the NRAO SCHED software, and the geodetic SKED program.
In addition to the Mark 5A recordings DiFX can also process input data recorded in Mark 5B format. A first implementation of a subclass of the new VDIF standard is at the test stage.
Correlator output is written according to the **FITS Interferometry Data Interchange Convention**. In addition to the fundamental visibility function measurements and associated meta-data, the FITS files include amplitude and phase calibration measurements (soon), and editing flags, that are logged at the VLBI antennas. AIPS release 31DEC08 or later is required to handle DiFX data properly.
MK4 formatted output is presently under test. The MK4 fringe-fitting program fourfit has been modified already by Capallo (MIT Haystack) and will allow a seamless path from the DiFX correlator into the geodetic analysis software. It will also allow astronomers to use the MK4 fringe-fitting for e.g. mm-VLBI or detection experiments.
The **Bonn DiFX implementation** is significantly more powerful and flexible than the previous MK IV hardware correlator, which is still used for some geodetic observations. Up to 15 stations have been correlated in one pass so far. 13 Mark 5 units are directly available for playback plus huge disk storage for holding data for correlation from more than 13 telescopes. Output datasets of several 100 GB can easily be handled.

Capabilities

DiFX can be configured to produce visibilities with pretty much arbitrary time and frequency resolution (although you need to be aware of the fundamental limitations of FFT window sizes etc.), and can produce an arbitrary combination of products on each baseline (i.e. full stokes on one baseline, parallel only on another etc.). It can channelize using an **FFT or a polyphase filterbank** (slower, but better channel spectral response - good for isolating RFI or spectral lines). Finally, it can **pulsar bin** visibilities into any number of bins, using **incoherent dedispersion**. This requires some a priori information on the pulsar ephemerides, provided in the form of a polyco file produced by the pulsar software package TEMPO. If you're just trying to recover the best possible SNR from a pulsar observation (e.g. to do astrometry) then DiFX gives you the option to sum your bins, after weighting each by the power contained in that bin. This gives the best possible SNR, by integrating over the whole pulse period.

Spectral Resolution

DiFX currently supports powers-of-2 numbers of spectral points spanning each individual baseband channel, up to **4096** for routine DiFX processing, and up to **32,768** if required and adequately justified. (The latter limit is the maximum resolution currently supported by AIPS.) **Oversampled data** (essential for extremely high spectral resolution with the existing VLBI baseband subsystems can be decimated appropriately. Currently, both the number of spectral points, and the oversampling factor, must be the same for all basebands at any given time, although multiple passes with different baseband subsets are possible. The actual spectral resolution obtained, and statistical independence of the spectral points, depends on subsequent smoothing and other processing.

A new DiFX capability supports **spectral zooming**, selection of a subset of correlated spectral channels from any or all basebands. Only the selected channels are included in the output dataset. This capability will be of value mainly in maser studies, where the recorded band may be much wider than the maser emission in two main categories of observations:

- Maser astrometry with in-beam continuum calibrators. Wideband observing is required for maximum sensitivity on the calibrators, while zooming allows high spectral resolution at the frequencies where maser emission appears.
- Multiple maser transitions. When wide bands are used to cover a large number of widely separated maser transitions, spectral zooming allows the empty portions of high-resolution spectrum to be discarded.

In proposing observations that will use spectral zooming, the required number of channels before zooming should be specified in the Proposal Submission Tool. Initially, the location and width of the "zoom" subbands will have to be communicated directly to the correlator before correlation.

Integration Period

DiFX accommodates a nearly continuous range of correlator integration periods over the range of practical interest. Individual integrations are quantized in multiples of the indivisible internal FFT interval, which is equal to the number of spectral points requested, divided by the baseband channel bandwidth. Since the latter are powers-of-2 MHz, the internal FFT interval is always a power-of-2 microseconds.

For most cases, with low to moderate spectral resolution, and/or wideband baseband channels, the FFT intervals are fairly short, and it is straightforward to find an integration period in any desired range that is an optimal integral multiple of the FFT interval, where optimal refers to the performance of DiFX. Extreme cases of very high spectral resolution (many spectral points across a narrow baseband channels – resolution of less than about 100 Hz) imply FFT intervals long enough that only limited choices of integral multiples are available.

For flexibility in these situations (although the option exists in all cases), integration periods other than an integral multiple of the FFT interval can be approximated, in a long-term mean, by an appropriate sequence of nearby optimal integral multiples. In this case, output records are time-tagged as if correlated with exactly the requested period.

Multiple Phase Centers

The field of view of a VLBI interferometer using typical correlation parameters (e.g., 0.5 MHz spectral resolution, 2 second averaging) is very small, around 1/10,000 of the primary beam area of the biggest telescope involved. Thus, to image targets which are widely spaced in the primary beam requires multiple processing passes in typical correlator implementations. If the visibilities are maintained at high time and frequency resolution, it is possible to perform a uv shift after correlation, essentially repointing the correlated dataset to a new phase center. However, this approach requires large visibility datasets.

A new feature in the DiFX correlator implements multiple uv shifts inside the correlator, to generate as many phase centers as are necessary, in a single correlation pass. The output consists of one dataset of normal size for each phase center. This mode consumes around three times the correlator resources of a normal continuum correlation, due to the need for finer frequency resolution before the uv shift, but the additional cost is only weakly dependent on the number of phase centers. For example, correlating 200 phase centers requires only 20% more correlator time than 2 phase centers. This mode thus should only be requested for imaging of three or more sources within any single antenna pointing.

This mode has to be requested and justified in the proposal and later communicated to the correlator in Bonn. Updates to the SCHED program, to be announced subsequently, will support specification of the actual phase center locations. The requested spectral resolution and integration time should correspond to the desired initial number of frequency channels per subband (required to minimize bandwidth smearing) and the desired integration between uv-shifts (to minimize time smearing).

Output Rate

Correlation parameters can result in an output rate of more than 10 MBytes per second of observing time; such high data rates must be adequately justified if required, as they will slow down the correlation. Observers should ensure that their data-analysis facilities can handle the dataset volumes that will result from the correlation parameters they specify.
Approximate output data rates are predicted by the SCHED software.

Experiments correlated in Bonn using DiFX since summer 2009

Code	# stations	Duration	Comments
C092A	12	5 x 24h	mm-VLBI
GB065A	15	15h	Global VLBI
EP066	6	2x24h	EVN, some sources recorrelated with high spectral resolution
EM077A	11	16h	EVN
GB065B	14	15h	Global VLBI
EM080A	6	12h	EVN pulsar experiment, correlated with and without gating. Recorrelated centered on pulsar coordinates obtained from the first correlation run
etc.			Several test & networking monitoring experiments

- References:
  - Adam Deller: Swinburne University Melbourne
  - Walter Brisken: NRAO (VLBA observational status)