

DiFX2.0

What you need to know

DiFX2.0 is more revolution than evolution. The final data products are the same (the correlator binary output format and the FITS files built from this are identical to 1.5) but the configuration files (.input, .calc etc) and the internals of mpifxcorr have been sliced, diced and reorganised. (See the [file format](#) page for the new DiFX2.0 file formats, and a link to the older 1.5 style formats). These changes are designed to enable the cool new features described below. However, no attempt has been made to keep DiFX2.0 backwards compatible with DiFX1.5. That is: you need to regenerate all the control files for a given experiment in order to run it. DiFX2.0 can't understand DiFX1.5 control files, and vice versa.

The user perspective: new features

DiFX2.0 lets you do a whole bunch of exciting new things - this is why we went to the trouble and heartache of breaking backwards compatibility. In no particular order, the new features are:

- [In-correlator spectral selection and averaging](#)
- [Multiple simultaneous phase centres](#)
- [Fast autocorrelator dumping for transient searches](#)
- [Configurable phase-cal tone extraction](#) *COMING SOON*
- [A phased array \(instead of cross-correlation\) output](#) *COMING SOON*
- [LO offsets](#) (post-F, so limited to ~kHz)
- [Arbitrary order clock models](#) (rather than just delay+rate)

Each of these new capabilities is discussed in more detail individually - click on them to get the full story

Under the hood: code changes for performance improvement

DiFX2.0 has had several important tweaks that improve efficiency on both the station-based and baseline-based side of processing. These are:

- Making use of linear approximations to substitute complex multiplications for much of the trigonometric operations in fringe rotation, fractional sample correction and (new in DiFX2.0) the phase shifts for multiple phase centre correlations. Click [here](#) for a more detailed explanation of the station-based improvements.
- Optimising baseline-based processing for the situation with large arrays and/or large numbers of spectral points, meaning the entire accumulator vector cannot remain in L2 cache. This is implemented by “batching” several FFTs in a row for one station, and then “batching” the cross-multiplication and accumulation. Click [here](#) for a more detailed explanation of the baseline-based improvements.

The station-based processing improvements lead to a ~25% reduction in the station-based cost of correlation (for time domain fringe rotation, which is almost universally used). For a 10 station array like the VLBA, which is dominated by the station-based processing, this means that DiFX2.0 is roughly 20% faster than DiFX1.5 for normal continuum correlations.

The baseline-based processing improvements offer no advantage for normal continuum correlations, but allow high spectral resolution correlations (required for e.g. multiple phase centres) to be undertaken with a much smaller penalty. For the VLBA, running a normal experiment with 1024 spectral points per subband instead of 128 resulted in a factor of ~5 slowdown - for DiFX2.0 this penalty is less than a factor of 2. Most of the remaining slowdown is actually due to the increased computational load of the larger FFT, and as such can't be avoided.

Finally, DiFX2.0 uses an all-binary data format for the visibilities and metadata that it dumps to disk, and it buffers in memory before writing for much better efficiency. DiFX2.0 can thus sustain much higher write speeds (on a standard single disk, ~50 MB/s vs ~5 MB/s for DiFX1.5. RAID disks would probably allow much higher rates.)

[Chris' Notes on Configuration Code](#)

From:

<https://www.atnf.csiro.au/vlbi/dokuwiki/> - **ATNF VLBI Wiki**

Permanent link:

<https://www.atnf.csiro.au/vlbi/dokuwiki/doku.php/difx/difx2.0doco>

Last update: **2015/10/21 10:08**

