SDP Memo 105: SDP Intermediate Data Products

Document number…………………………………………………………...SKA-TEL-SDP-0000210
Document Type………………………………………………………………………..... MEMO
Revision……………………………………………………………………….....................01
Author……………………………………………………………………………… Anna Scaife
Release Date………………………………………………………….................. 2019-03-27
Document Classification……………………………………………………................. Unrestricted

Lead Author	Designation	Affiliation
Anna Scaife		
Signature & Date:		

SDP Memo Disclaimer

The SDP memos are designed to allow the quick recording of investigations and research done by members of the SDP. They are also designed to raise questions about parts of the SDP design or SDP process. The contents of a memo may be the opinion of the author, not the whole of the SDP.

Introduction

This memo is taken from a Confluence page written by Anna Scaife: https://confluence.ska-sdp.org/display/SD/Intermediate+Data+Products.

Data Driven Architecture (DDA)

Required (sky derived) gain tables:

- K (delay - N_ant),
- G (time dependent instrumental gain - N_int x N_ant x 2),
- B (frequency dependent bandpass - N_f x N_ant x 2),
- XYf (frequency dependent cross-hand phase - N_f x N_ant),
- D (frequency & time dependent polarization leakage - N_int x N_ant x N_f),
- Z (time dependent ionospheric phase - N_int x N_ant),
- F (time dependent Faraday rotation - N_int x N_ant)

Z & F can have frequency dependence calculated analytically, maybe no need to store - but might be required for F in the case of complex Faraday depth spectra.
If multi-direction calibration is performed (required for LOW) then G, Z & F have an additional dimension of N_dir (or N_facet). RJN: D as well. For LOW the polarization response is direction dependent. And actually, I guess that would be the case for MID as well, if you go to high dynamic range (due to, e.g., pointing errors).
Intermediate imaging products:

- Residual Image (N_pix x (N_tt+1) x N_scale),
- Component List (N_comp x (N_tt+1)),
- Dirty Image (N_pix x (N_tt+1) x N_scale),
- PSF Image (N_pix x N_scale)

Local Telescope Model (LTM)

The LTM will need to contain all the information required for the imaging and calibration functions, or data that these information can be derived from. Some of these data will need to be updated as a function of time, but this can be done in a predictive fashion for most things if flagging for error states has already been applied. An initial (non-exhaustive) list comprises:

- Jones matrices for the antenna response (X,Y)
- Parallactic angle, or (preferred):

    - longitude of telescope
    - latitude of telescope
    - RA of phase centre
    - Declination of phase centre
- Antenna positions in some defined co-ordinate system
- Flags on antennas (missing etc) - if not already applied ?
- Elevation (antenna based)
- Azimuth (antenna based)
- Reference frequency
- Reference frequency channel
- Frequency channel width/increment
- Number of frequency channels
- Polarizations (and ordering)
- Time (in UTC?) or MJD reference
- Time increment/step

Time dependent elevation & azimuth may not be required if flags can be generated before SDP for pointing issues, although elevation is probably needed at the higher frequencies for air mass corrections? (ask PIP.CAL) This could also be calculated from (long, lat, RA, Dec, time).

The copy of the LTM within SDP will need to be updated every time an averaging operation is performed in time/frequency, or a phase rotation operation is performed. Otherwise the LTM can be extracted from TM only once at the start of each observation, so on a cadence of ~6 hours.

Beam Models

Beam models are required for calibration (where they are used to adjust the LSM) and for imaging, where they are used in the A-projection and mosaicing steps. Beam models will need to be held as Jones matrices for calibration as well as for imaging, where gridding is likely to be done in (XX, XY, YX, YY). If possible it would be best to hold these beam models as a set of basis functions in order to minimise storage requirements, although this is not yet done for most software. (Need to check for LOFARsoft).

Note: If uvw coordinates are not passed directly from CSP to SDP, but need to be recalculated within SDP, then the LTM will require a significantly larger amount of information about the array. The SDP will also require SOFA as part of CAS.

Local Sky Model (LSM)

The GSM/LSM is discussed in more detail o this Confluence page: https://confluence.ska-sdp.org/display/SD/2.+GSM+Issues, including details of cadence of updates and format/method of updates. Given the changing astrophysics across the different SKA1 bands there will probably need to be a distinct LSM per band. These will need to contain the following information:

- Position of component (including uncertainties);

- Stokes I Taylor terms (tt0, tt1, tt2) of component (including uncertainties);
- Stokes Q tt0 component;
- Stokes U tt0 component;
- Stokes V Taylor terms (tt0, tt1, tt2) of component (including uncertainties);
- Rotation Measure (if known);
- Reference frequency for tt0 / RM;
- Form of component (i.e. point, Gaussian, wavelet etc);
- Parameters of component (including uncertainties);
- Variability flag;
- Spectral line flag(s) ?
- A unique identifier, including a term identifying associated components.

Note: Q, U, V values are not required for calibration but are required for subtraction of the LSM from the visibility data in the Slow Transients (Fast Imaging) Pipeline (STP).

Should the LSM be a gridded model or a component list?

In the most extreme case (fully filled FoV), these two things are equivalent. For the STP, it minimises FLOPs just to consider a component list rather than a gridded model. Indeed most (if not all - I can't think of an example) existing telescopes use component list based LSMs. (It's possible to use an image in CASA, but in practice I'm not sure it's worth it because the image is usually just made from a component list in a prior step). We will use the parametric model to evaluate the FLOP trade-off for realistic source populations.

How much LSM Management is done with component lists and how much is done with actual sky images?
For imaging, all LSM management is done with component lists.
For calibration, it is possible to use an image - but I don't know what the PIP.CAL people think about this.
For updating the GSM based on LSM evaluation, component lists are much simpler to handle.

How are sources represented (point source components, shapeless or some other compact basis representation etc.)?
The MS-CLEAN uses point sources and Gaussians of varying width, so updates to the LSM/GSM from the imaging pipeline will naturally occur in this format. Wavelet-based CLEAN was implemented for Optical & MIR data in the 90s in a way that's basically the same as MS-CLEAN but with wavelet planes rather than Gaussian scales, so it could be done without any change to the architecture. Compressed Sensing reconstructions use a variety of wavelet bases, but it's difficult to justify a particularly optimal basis for either CLEAN or CS.

It should not be required that the only component representation is that supplied by the imaging pipeline. More complex representations should be included for well-studied complex objects that have been derived from external analyses.

How is frequency structure represented?

The standard Taylor term expansion seems a sensible way to go for Stokes I and probably Stokes V also. Stokes Q and U are not well represented by this expansion due to Faraday rotation, but could be defined by a rotation measure and Q,U values at some reference frequency. These QUV data are not required for calibration but only for frequency dependent subtraction in the STP.

Are only continuum sources in the LSM?
Yes, I think so. Spectral line information is not required for calibration (even in the Spectral Imaging pipeline), but the effect of spectral lines ion the STP should be assessed in case they are misinterpreted as transient sources (although it should be easy to discriminate in the source finding).

Are standard calibrator source models stored in the LSM and if so how are these data queried?
Yes. Components for specific sources can be extracted using their unique identifiers, including associations of components that form a coherent source structure. Most source finding algorithms do this kind of grouping automatically, but these models can also be refined externally to the standard pipeline and then updated.

If shapelets or an equivalent are used can we be generic at this a=stage or do we need to be specific (e.g. sources represented in terms of a compact basis set representation ...?).
We can probably afford to be generic and let this evolve over the lifetime of the telescope. It should be possible to add in different bases later.

Note that the baseline architecture should assume that clean-component models generated in the main imaging loopier treated as data (see above to document) and not as part of the LSM. But the output should be assessed relative to the input LSM in order to check for updates.