Files and resources to generate the figure 1 in Bowman et al.(2018)

Steven Murray, Nivedita Mahesh, Alan Rogers
ASU

03/30/2022

Introduction

The purpose of this brief memo is set in stone the code, parameters and input data required to (re)produce the H2 case from Bowman et al. 2018 (hereafter B18), which also corresponds to Fig. 1 of that paper.

These codes and scripts were provided by Alan, and adapted by Steven and Nivedita to i) ensure that they work wherever they are called from (as long as it's on enterprise) and ii) to prepare them for being stored in a GitHub repository for longer-term reproducibility and access.

The main code

We have created a github repository for Alan's pipeline at `https://github.com/edges-collab/alans-pipeline` (it is private, at least for now). While some of the codes in it have been very minimally modified with respect to the original codes that Alan provided, this modification is restricted to either (i) extra printing either to screen or file, for the purpose of collecting more data to compare to other pipelines, or (ii) edits made to bring the code *more* in line with the original code Alan actually used for these results (the code he provided had already been modified by him compared to the code run back in 2017/2018).

Note also that while this repository may evolve into the future, the results mentioned in this memo can always be exactly re-obtained by using the `nature-paper-case-h2` tag of the repo; i.e. by doing `git checkout nature-paper-case-h2` in the repo before installing/running.

In this code, there are three main executables:

1. `acqplot7amoon` – reads the raw `acq` files and performs filtering and averaging/smoothing. This is used both for lab-based spectra (to get lab calibration) and field data (prior to calibration).
2. `edges2k` – can either perform lab calibration or calibrate and perform corrections (beam, loss, etc.) on each field data file.
3. `longav` – averages over days and fits for the foreground and signal.

To be extremely clear, the Nature Paper H2 results (i.e. public results) can be obtained with the following commands:

```
git clone git@github.com:edges-collab/alans-pipeline
cd alans-pipeline
git checkout nature-paper-case-h2
make
cd scripts
csh run-H2-cal
csh run-H2-field-data
```

Settings in the code

To obtain the results for the H2 case (i.e. that shown in Fig. 1 of B18), two steps need to be taken: first the lab calibration, then the reduction of the field data. Both of these steps can be performed exactly by running scripts in the `alans-pipeline` repository, specifically: `scripts/run-H2-cal` for the lab calibration, and `scripts/run-H2-field-data` for the field data reduction. Both of these scripts can simply be run with `csh <script>`.

Name	Description	Value
`-smooth`	Number of freq bins smoothed over with gaussian	8
`-fstart`	Minimum frequency to read	40 MHz
`-fstop`	Maximum frequency to read	110 MHz
`-pfit`	Number of terms to use in a polyfit in order to flag RFI	27
`-rfi`	Number of sigma away from model to flag as RFI	10
`-nrfi`	Number of bins on either side of channel to flag as RFI	0
`-tcal`	Initial guess for temperature of internal load + noise source	1000 K
`-tstart`	The hour at which to start reading data for the day	0 (all)
`-tstop`	The hour at which to stop reading data for the day	23 (all)
`-delaystart`	Number of seconds at start of observation to exclude	7200

Table 1: Options passed specifically to `acqplot7amoon` for the lab calibration spectra.

While this alone should be enough to fully reproduce the data, we here note all of the settings used in each of these scripts, *just in case*. Note that we're only specifying the non-default settings here; in case the underlying C code changes defaults in the future, you should always check out the exact git tag corresponding to this memo.

Lab Calibration – `run-H2-cal`

To do lab calibration, the script first reads each of the load spectra: ambient, hot, open and short, averaging them over time, and smoothing them over frequency, using `acqplot7amoon`. The input parameters used for *all files* are given in Table 1. Following this, the output averaged spectra are passed to `edges2k` to fit the noise-wave parameters. The parameters to this code are given in table 2.

The primary output of this script is the file `specal.txt`, which is used by the next script. The script also copies this file to the location `specal_<DATE>.txt` so that you can maintain a versioned copy. The specific `specal.txt` that corresponds exactly to the H2 case in B18 is included directly in the repository at `results/specal_final_case2.txt`.

Field Data Reduction – `run-H2-field-data`

The second processing script uses `acqplot7amoon` to read and reduce many days worth of field data (exact days given in next section), then applies the output `specal.txt` from the previous script to each day with `edges2k`, then finally averages all the days together and performs a model fit with `longav`. The exact settings for each of these three are given in tables 3, 4 and 5 respectively.

The primary output files of this script are the following (if using the script in the repository, all these files will be saved to the `H2Case/` directory):

1. `spegva<day>.txt` – these are the output of `acqplot7amoon`, which is called once per day. The contents are the spectra averaged over time (between the GHA limits given)

2. `spesum2_<date>.txt` – this is the calibrated spectra from each day, all concatenated together into a file; the output of `edges2k` (which is called once per day, taking in `spegva<day>.txt`).

3. `final_average_<date>.txt` – the final averaged spectrum processed through `longav` (which takes in the `spesum2.txt`). This output should exactly correspond to the final public results online, or more specifically, to `results/H2-final-average.csv` in the repository.

Files required

Along with the input settings, numerous files are required as input data. Some of these files are directly produced by the previous script, and we don't mention those here. Some of the files are specified on the command-line, and some are hard-coded into the C code. We list all such files here.

Files included in the repository:

1. `data/newniv.txt` – the beam model used for the low-band. It is copied to the script directory by the script that uses it (`edges2k` when applied to field data).

Name	Description	Value
-fstart	Minimum frequency of S11 files	40 MHz
-mfit	Number of parameters in fit (**Unused here**)	1
-wfstart	Minimum frequency of spectra	50 MHz
-wfstop	Maximum frequency of spectra	100 MHz
-cfit	Number of poly terms for scale/offset	6
-wfit	Number of poly terms for noise waves	5
-tcold	Assumed temperature of the ambient input	296
-thot	Assumed temperature of the hot load input	399
-nfit1	Number of Fourier terms for specal (**Unused here**)	27
-nfit2	Number of Fourier terms for fitting S11 of calibration loads	27
-nfit3	Number of Fourier terms for fitting S11 of LNA	11
-nfit4	Number of Fourier terms for fitting S11 of antenna (**Unused here**)	27
-Lh	-2 to use S11, S12 to get loss of hot load	-2
-lmode	2 to calculate loss of semi-rigid cable	2
-tant	Assumed temperature of 'antenna' (in this case, ambient temp for hot load loss calc)	296 K
-smooth	Number of freq bins to smooth over (**Unused here**)	8
-wtmode	Kind of weighting. Zero is standard.	0
-delaylna	An extra delay added to LNA from adapter, in seconds	0.0
-ldb	Extra attenuation on the LNA, in db	0
-delaycorr	An extra delay added to LNA from VNA, in seconds	0.0
-lna_poly	If 0, forces the LNA model to be a Fourier series	0

Table 2: Input Options to `edges2k` for producing the lab calibration (output is `specal.txt`). Note that `lna_poly` is a parameter added specifically to the code adapted for the repository, and not present in Alan's code. It allows forcing the LNA model to be a Fourier series, regardless of number of terms (since Alan's original code chooses it based on number of terms, but the threshold number of terms has changed since 2017).

Name	Description	Value
-smooth	Number of freq bins smoothed over with gaussian	8
-fstart	Minimum frequency to read	40 MHz
-fstop	Maximum frequency to read	100 MHz
-pfit	Number of terms to use in a polyfit in order to flag RFI	37
-rfi	Number of sigma away from model to flag as RFI	2.5
-nrfi	Number of bins on either side of channel to flag as RFI	4
-tcal	Initial guess for temperature of internal load + noise source	1000 K
-gha	Central GHA to include in spectral data.	12
-dgha	Width of included GHA on each side of `-gha`	6
-pkpwrm	Maximum power allowed above 80 MHz (for roughly calibrated temperature) to keep entire integration	40 K
-peakpwr	Maximum percentage of power above 100 MHz compared to entire spectrum to keep integration	3%
-minpwr	Minimum percentage of power above 100 MHz compared to entire spectrum to keep integration	0.7%
-dloadmax	**Unused in the code**	1000
-adcov	Max ADC value to accept spectrum	0.4
-maxrmsf	Max RMS of power-law fit with index -2.5 between 60-80 MHz to keep integration	200
-maxfm	Maximum difference between a channel and the average of its next-neighbours, between 88-120 MHz to keep integration	200 K

Table 3: Input options to `acqplot7amoon` for the field spectra.

Name	Description	Value
`-fstart`	Minimum frequency of S11 files	50 MHz
`-fstop`	Maximum frequency of S11 files	100 MHz
`-mfit`	Number of parameters in spectrum fit (affects smoothing)	3
`-wfstart`	Minimum frequency of fit	51 MHz
`-wfstop`	Maximum frequency of fit	99 MHz
`-nfit4`	Number of Fourier terms for fitting S11 of antenna (why?	10
`-lmode`	6 to calculate loss of balun and connector	6
`-tant`	Assumed temperature of antenna to calculate loss	296 K
`-smooth`	Number of freq bins to smooth over before fitting	8
`-wtmode`	Kind of weighting. Zero is standard.	100
`-delaylna`	An extra delay added to LNA from adapter, in seconds	0.0
`-ldb`	Extra attenuation on the LNA, in db	0
`-delayant`	An extra delay added to Antenna from adapter, in seconds	0.0
`-adb`	Extra attenuation on the Antenna, in db	0
`-skymode`	Along with `-site` and `-map`, sets whether to do beam correction, and what sky to use	384 (Haslam at MRO)
`-antaz`	Rotation of antenna used for beam model	354°
`-low`	**Unused here**	1
`-mdd`	Chooses which FG model to use	4 (linphys)
`-cmb`	How to deal with the CMB in beamcorr. "2" means add CMB to sky model, but make no other correction	2
`-eorcen`	Guess of central EoR frequency	78 MHz
`-eoramp`	Guess of EoR amplitude	-0.5 K
`-eorwid`	Guess of EoR width	0 MHz
`-tau`	Guess of EoR τ parameter	7

Table 4: Input Options to `edges2k` for calibrating field data.

Name	Description	Value
`-fstart`	Minimum frequency to fit	51 MHz
`-fstop`	Maximum frequency to fit	99 MHz
`-lim`	Maximum RMS of the FG model to the spectra to include the day	0.17 K
`-nfit`	Number of terms in FG fit (5 to do linphys)	5
`-out`	Specifies what columns to write in the output file	3
`-dmax`	**Sets scale for plots only**	0.5
`-schk`	Flag GHA based on sun position if > 0	0
`-tchk`	Flag GHA if estimated T_{75} is lower than this	200 K
`-rfi`	Flag frequency if diff with the next is larger than the RMS over freq times this number	1.9
`-g10`	Used for plots only	True
`-date`	Just specifies how to read input file dates	True
`-seor`	If negative, get SNR of EoR signal	-1
`-tau`	Fix τ in EoR model fits to this	7
`-sig`	specify which kind of signal model to fit	0 (flattened gaussian)
`-md`	Whether to use flattened Gaussian in SNR estimate	1

Table 5: Input Options to `longav` for averaging field spectra over days.

2. `data/S11_blade_low_band_2015_342_03_14.txt.csv` – The calibrated antenna S11 measurement (calibrated by Raul, provenance not exactly known).

3. `data/s11_calibration_low_band_LNA25degC_2015-09-16-12-30-29_simulator2_long.txt` – the calibrated (but unmodelled) S11's of the LNA and lab-calibration loads. Calibrated by Raul using the AGILENT 85033E calkit, resistance of 49.98Ω for the LNA and 50.12Ω for the internal switch, and 30 ps delay for the match standard (for the LNA). Input uncalibrated LNA measurements were taken in 2015 on days (08-28, 08-31, 09-01, 09-02, 09-03, 09-05), and averaged together after calibration.

4. `scripts/408-all-noh` – The Haslam sky map used for getting the beam correction.

Files not included in the repository (present on enterprise in `data5/edges/data/`):

1. **Input calibration spectra** – All located in `CalibrationObservations/Receiver01/Receiver01_25C_2015_09_02_040_to_200MHz/Spectra`. Specific files are:
   - `Ambient_01_2015_245_02_00_00_lab.acq`
   - `Ambient_01_2015_246_02_00_00_lab.acq`
   - `HotLoad_01_2015_246_04_00_00_lab.acq`
   - `HotLoad_01_2015_247_00_00_00_lab.acq`
   - `LongCableOpen_01_2015_243_14_00_00_lab.acq`
   - `LongCableOpen_01_2015_244_00_00_00_lab.acq`
   - `LongCableOpen_01_2015_245_00_00_00_lab.acq`
   - `LongCableShorted_01_2015_241_04_00_00_lab.acq`
   - `LongCableShorted_01_2015_242_00_00_00_lab.acq`
   - `LongCableShorted_01_2015_243_00_00_00_lab.acq`

2. **Input Field Spectra** – All located in `2014_February_Boolardy/mro/low/<year>` where `year` is either 2016 or 2017. No files are skipped. The first day to be used is 2016-250. The last day to be used is 2017-094.

Comparing the spectra produced with the publicly available data

The first point of reference we have is the calibration file, produced by `edges2k`. We have the original file that was produced by Alan (in the `results/` directory of the repo). We note that **comparing our output to this calibration file, we get *exact* correspondence**.

Going further, to confirm that the script has generated the same spectra as the data that was used in Fig. 1 of B18, we plot the residuals to a 5-term LinLog foreground model in Fig. 1. We plot both the output data from `alans-pipeline` (orange dashed) and the directly-downloaded data from the LoCo website (thick black). They ***exactly correspond***.

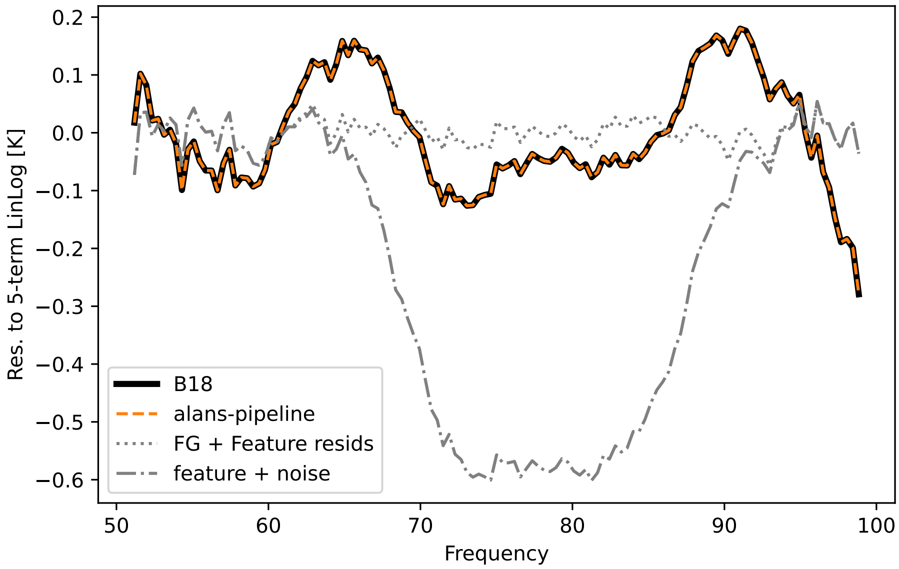

Figure 1: Comparison of Nature Paper data (from Fig. 1 of B18) to the spectrum obtained with the above processing pipeline. Both are shown as residuals to a 5-term linlog. The difference between the two is exactly zero. Also included are the residuals to a 5-term linlog and flattend-gaussian absorption feature (dotted gray), and the fitted absorption feature plus noise (gray dot-dash).