The Data processing pipeline for the GUSTO mission

Russell Shipman,[1,2] Youngmin Seo,[3] Volker Tolls,[4] William Peters,[5] Ümit Kavak,[2,1] Craig Kulesa,[5] and Chris Walker[5]

[1]*SRON, Groningen, Groningen, The Netherlands;* `R.F.Shipman@sron.nl`

[2]*Kapteyn Astronomical Institute, Groningen, The Netherlands*

[3]*JPL Caltech, Pasadena, California, USA*

[4]*Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA*

[5]*University of Arizona, Tuscon, Arizona, USA*

Abstract. The Galactic Ultra-long duration Stratospheric Terahertz Observatory (GUSTO) is a stratospheric balloon mission to map 124 square degrees of the Galactic Plan and LMC in the 3 brightest interstellar cooling lines: [CII], [OI], and [NII] at 158, 63, and 205 microns, respectively. The mission will use heterodyne mixers to velocity resolve the emission lines and is planned to be launched in December of 2021. The GUSTO data processing pipeline will produce baseline corrected data cubes available to the general community. This poster will present the approaches and challenges to be used to processes the data.

1. Introduction

Observing abundant atoms and ions traces the dynamics and life cycle of the interstellar medium (ISM) in galaxies. Singly ionised carbon, [CII], is particularly useful to trace the ISM. With an ionisation potential of 11.26eV, carbon is easily ionised by the UV radiation from young hot stars. High spatial and spectral observations of [CII] show not only where the ion is but how it is moving. Sparse spatial but velocity resolved [CII] emission has recently been studied along 500 lines-of-sight throughout our Galaxy (Langer et al. 2014) and provided significant insights into the internal workings of the ISM in our Galaxy.

OI and NII

GUSTO will use 3 arrays of 8 HEB mixers. The CII and NII LO's are provided by JPL and the OI LO by a Quantum Cascade Laser (QCL) which will be Fourier multiplied to produce 8 LO signals.

GUSTO will observer CII, OI and NII fine structure lines for 124 square degrees of the Galactic Plane and the Large Magellanic Cloud.

Details of the STO experiment can be found in Walker et al. (2010).

2. Observations

For the [CII] survey, GUSTO will make use of the On-The-Fly mapping (OTF) mode. OTF is a means of mapping a region by continuously scanning and intermittently reading out a detector (Mangum et al. 2000). OTF is a highly efficient means of covering a large region of the sky with single detectors or small arrays of detectors. The OTF technique uses the standard vane calibration of radio telescopes (Kutner & Ulich 1981) which combines data of a sky reference position free of emission as well as data on an internal load of known temperature. STO-2 has an internal hot load. The main constraint in OTF mapping is the timing between readouts of the instrument during the scan (ON), the time between load measurements (HOT) and the time between the sky measurements (REF).

An OTF scan is shown in Figure **??**. The scan begins with a sky (REF 1) observation along with a hot load (HOT). The telescope points to the beginning of the mapping region and repeats the load measurement (HOT_B). Then starts integrating on the sky while moving along the scan leg. The integrations are readout frequently to minimise source blurring. At appropriate intervals and at the end of the scan the internal hot load is observed (HOT_E). This pattern continues on a new scan parallel but offset by a fraction of the spatial resolution of the instrument.

As can be seen in the Figure **??**, the timing between successive ON readouts is the fastest, followed by the intermittent HOT, then by the REF measurements. This pattern comprises the observation cadence.

2.1. Standard calibration

The reference observations (REF) and repeated hot measurements (HOT) are combined to calibrate the system onto a known radiometric scale (Kutner & Ulich 1981). Reference positions should be emission free and the system should be stable. The calibration of a DBS receiver is give by:

$$\mathrm{T}_A^* = \mathrm{T}_{sys}\frac{(ON - REF)}{REF} \tag{1}$$

$$\text{with, } \mathrm{T}_{sys} = 2 \times \frac{\mathrm{T}_{HOT} - Y \times \mathrm{T}_{REF}}{Y - 1} \text{ with } Y = \frac{HOT}{REF}$$

T_{sys} is the system noise temperature. The Y factor is the ratio of the raw HOT counts to the raw REF counts. T_{HOT} is the hot load temperature (290K) while T_{REF} is the effective temperature of the blank sky at 1.9 THz (45K). The values of the Y factor and the REF measurement per channel are linearly interpolated to the time of the ON readout.

2.2. Radiometric noise and drift noise

Radio observations are afflicted with mainly two different noise types: radiometric (white) noise and drift noise. White noise is independent of frequency of observing and can be reduced by longer integrations. Drift noise increases over longer time periods and in general cannot be reduced by longer integrations or repeated measurements.

Drift noise originates from the instability of the detector system. Understanding the timing of instabilities is needed and a proper observation sequence must be chosen to minimise drift effects. Whereas white noise is flat across spectral channels (bandpass), drifts result in spectra which fluctuate over the bandpass. The resulting spectra suffer from poor baselines and/or standing waves. Poor baselines limit the useful information present in the signal by confusing spectral features of the sky with drift noise.

2.3. Addressing drift noise

The calibration requires stability of the entire system. The reference observations are usually taken a significant time before and after the OTF scan. The drift time constant is described by the Allan time (Allan 1966) and observations should be designed with this drift time in mind.

The frequency of the load observations (HOT) helps make up the overall cadence. Another component is the reference observation. Often, the stability time is short compared to the cadence of the reference measurements. This implies that the references observations, although necessary for calibration, may leave larger than desired drifts. The hot load can be used to address the system drift since changes in the HOT reflect the changes of the system much closer in time. In other words, an intermittent load scan can be used to help stabilise the calibration.

To account for drifts and better use the system monitoring aspect of frequent load measurements the calibration equation can be altered to create an "interpolated" REF signal.

$$\mathrm{T}_A^* = \mathrm{T}_{sys} \frac{(ON - intREF)}{intREF} \tag{2}$$

In this case, $intREF = HOT(t) \times \frac{REF(t_0)}{HOT(t_0)}$ and is linearly interpolated to the scan readout time t. t_0 is the time of the reference scan and accompanying hot.

Interpolating between standards is not new and a full discussion for OTF observations is given in Ossenkopf (2009). In the presence of significant system drifts, all calibration factors need to be interpolated in time to match the scan integration time.

3. Machine learning as a pipeline step

Instrument drifts will likely be present for GUSTO as that is in the nature of the HEB mixer. We will therefore have to make use of machine learning of the possible patterns which will be introduced. The patterns change over time, but have a lot of continuity. HIFI also had pattern matching corrections in its pipeline for the HEB mixers (**?**).

In order to properly introduce a ML step in the pipeline, we will need training and validation data. One approach to obtain the data sets is to observe blank fields. This is however a costly step. Another approach is to simulate standing waves. Such an approach was successful for STO data

4. Conclusions

The GUSTO pipeline will automatically identify spikes in the data, apply calibration, correct for detector drifts, remove standing waves and make data cubes (position - position, velocity).

These data will be made public for use by the astronomical community.

References

Allan, D. W. 1966, IEEE Proceedings, 54

Kester, D., Avruch, I., & Teyssier, D. 2014, in Bayesian Inference and Maximum Entropy Methods in Science and Engineering, vol. 1636 of American Institute of Physics Conference Series, 62

Kutner, M. L., & Ulich, B. L. 1981, ApJ, 250, 341

Langer, W. D., Velusamy, T., Pineda, J. L., Willacy, K., & Goldsmith, P. F. 2014, A&A, 561, A122. 1312.3320

Mangum, J., Emerson, D., & Greisen, E. 2000, in Imaging at Radio through Submillimeter Wavelengths, edited by J. G. Mangum, & S. J. E. Radford, vol. 217 of Astronomical Society of the Pacific Conference Series, 179

Ossenkopf, V. 2009, A&A, 495, 677. 0901.2486

Walker, C., Kulesa, C., Bernasconi, P., Eaton, H., Rolander, N., Groppi, C., Kloosterman, J., Cottam, T., Lesser, D., Martin, C., Stark, A., Neufeld, D., Lisse, C., Hollenbach, D., Kawamura, J., Goldsmith, P., Langer, W., Yorke, H., Sterne, J., Skalare, A., Mehdi, I., Weinreb, S., Kooi, J., Stutzki, J., Graf, U., Brasse, M., Honingh, C., Simon, R., Akyilmaz, M., Puetz, P., & Wolfire, M. 2010, in Ground-based and Airborne Telescopes III, vol. 7733 of Proceedings of the SPIE, 77330N