Building a prototype and testing system for processing GIFTS data

Maciek Smuga-Otto, SSEC
Ray Garcia, Bob Knuteson, Erik Olson, Jason Otkin, David Tobin, CIMSS

The Geosynchronous Imaging Fourier Transform Spectrometer (GIFTS) instrument will combine high spectral resolution soundings associated with Fourier Transform Spectrometers (FTS) with high spatial and temporal resolution, creating three-dimensional near-real time views of atmospheric radiance, temperature, water vapor, and winds.

GIFTS data processing: Background, Requirements, Architecture

Requirements for data processing software

* High Throughput: process 1.5 Terabytes of data per day.
* Low Latency: generate critical products within 5 minutes of gathering observation.
* Flexibility: Allow for easy development, testing and staging of new processing algorithms.
* Longevity: Software will evolve over a period of years to decades.
* Reproducibility: Record detailed processing history of data.
* Low Cost: Use off-the-shelf cluster hardware and leverage existing software technologies where possible.

The Data Processing Pipeline

Stream of incoming observations (interferograms) together with metadata

Separate by band, and optionally by scan direction and pixel

Initial FFT (go to spectrum) Stage

Complex Spectra

Data path: Interferograms

Create context from incoming observations

Metadata path: Retrieve records for time, band, scan direction, and (optionally) pixel index of incoming observations

Complex Spectra

Nonlinearity Study Stage

Use context to apply correction

Use context from blackbody observations

Radiometric Calibration Stage

Use context to calibrate

Real calibrated Spectra

Finite Field of View Correction Stage

Real calibrated Spectra

Spectral Resampling Stage

Real calibrated Spectra

Bundle reference (audit) metadata with outgoing calibrated spectra

The Cluster Environment

7 node research cluster at CIMSS

Front End Processor

Input Delivery

Cluster for processing L0-L1 Pipeline

Cluster for processing L1-L2 Pipeline

Data Bus

Data Archive and Distribution

Cluster processing of Winds

Mosaic Generation

Reshape Data Cubes

System Design Snapshot

Design of software layout at a single compute node of the cluster

Graph of logical dependencies between all software components of system