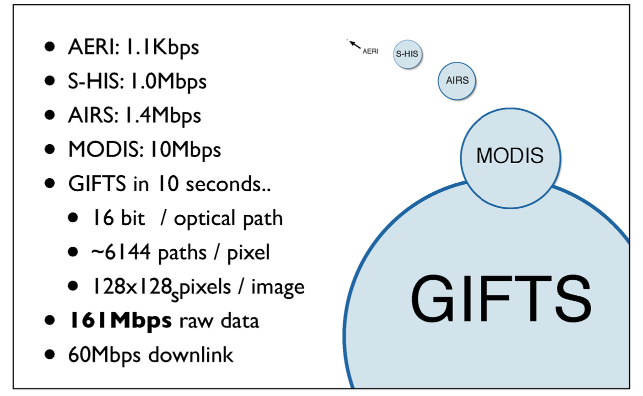
Distributed Computing for the Extraction of Meteorological Products from the GIFTS Imaging Interferometer

Raymond K. Garcia, Maciej J. Smuga-Otto, Hung-Lung Huang, Paolo Antonelli

Cooperative Institute for Meteorological Satellite Studies (CIMSS)
University of Wisconsin-Madison Madison, WI 53706, U.S.A.

We present an approach for processing the large volumes of data from an Imaging Fourier Transform Spectrometer (IFTS). Methodologies and technologies are shown for near-real-time extraction of products (radiance spectra, temperature and water vapor retrievals, multi-level wind vector fields). These techniques were originally adopted to cope with the expected data volume for the GIFTS instrument, and are directly applicable to an IFTS-based HES.



Challenges for data system:
- Large data volumes
- High data rates
- Low latency requirements
- Flexibility
- Management of reference data

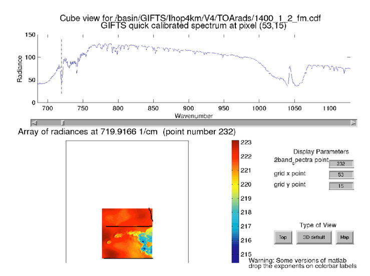
7 node research cluster at CIMSS

Definitions:

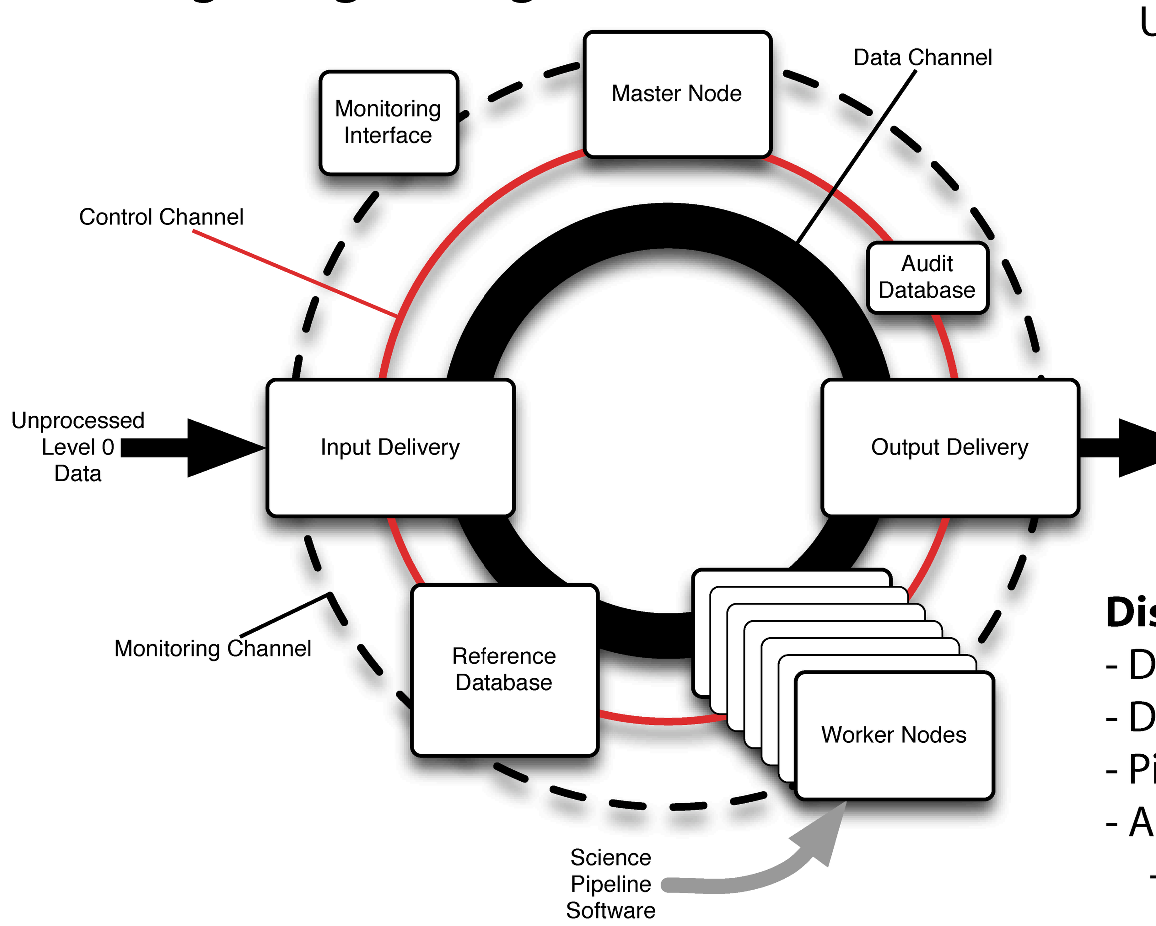
Pixel - a single infrared detector on the focal plane NxN detector array

Data Cube - NxN square array of data from single observation by IFTS
- Interferograms, Spectra, or retrieved T/WV
- 128x128 on GIFTS instrument (16384 spectra per observation)

Pipeline - Sequence of processing algorithms (Stages) applied to data from a single pixel



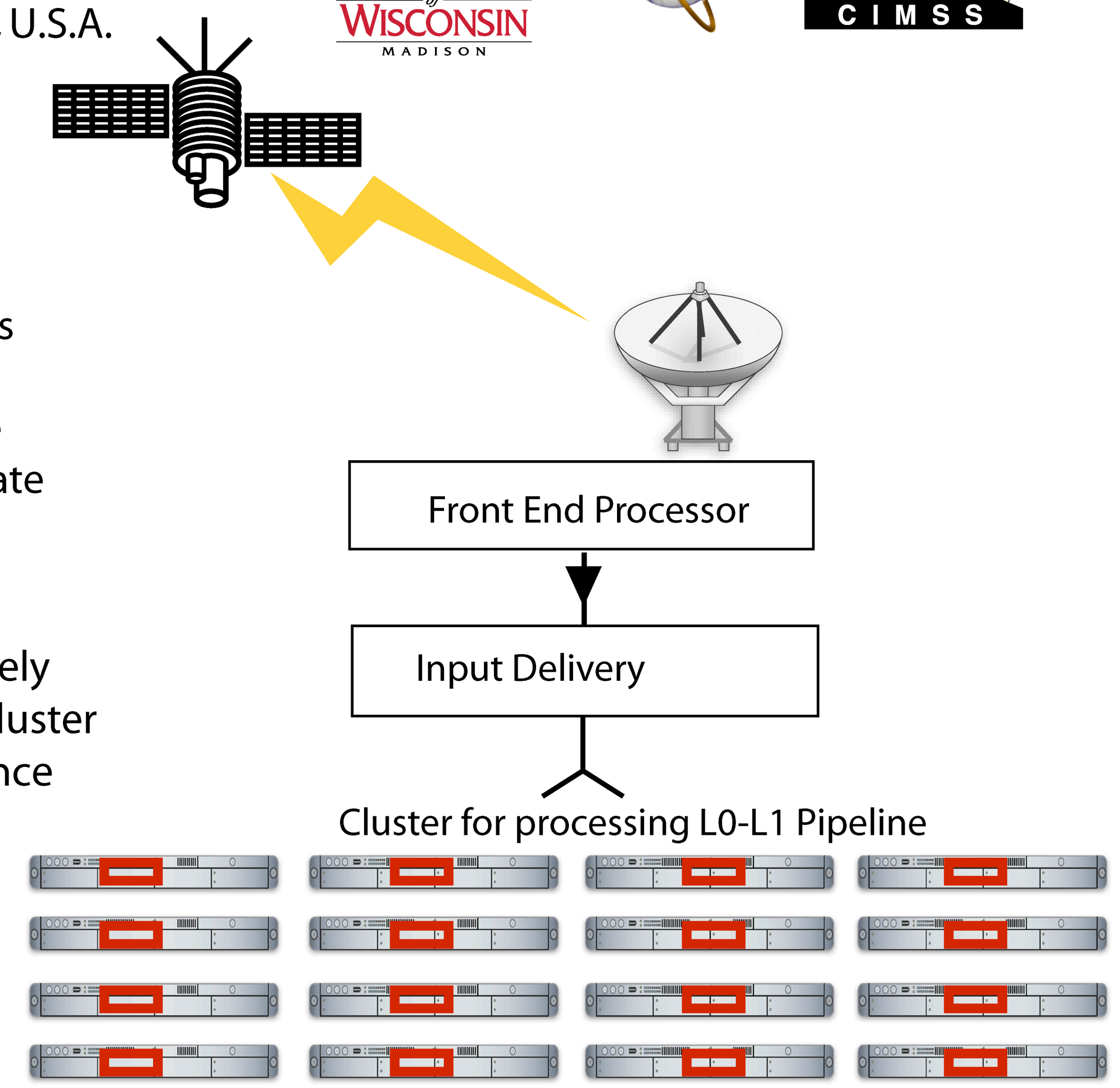
L1 Product: Calibrated Spectra

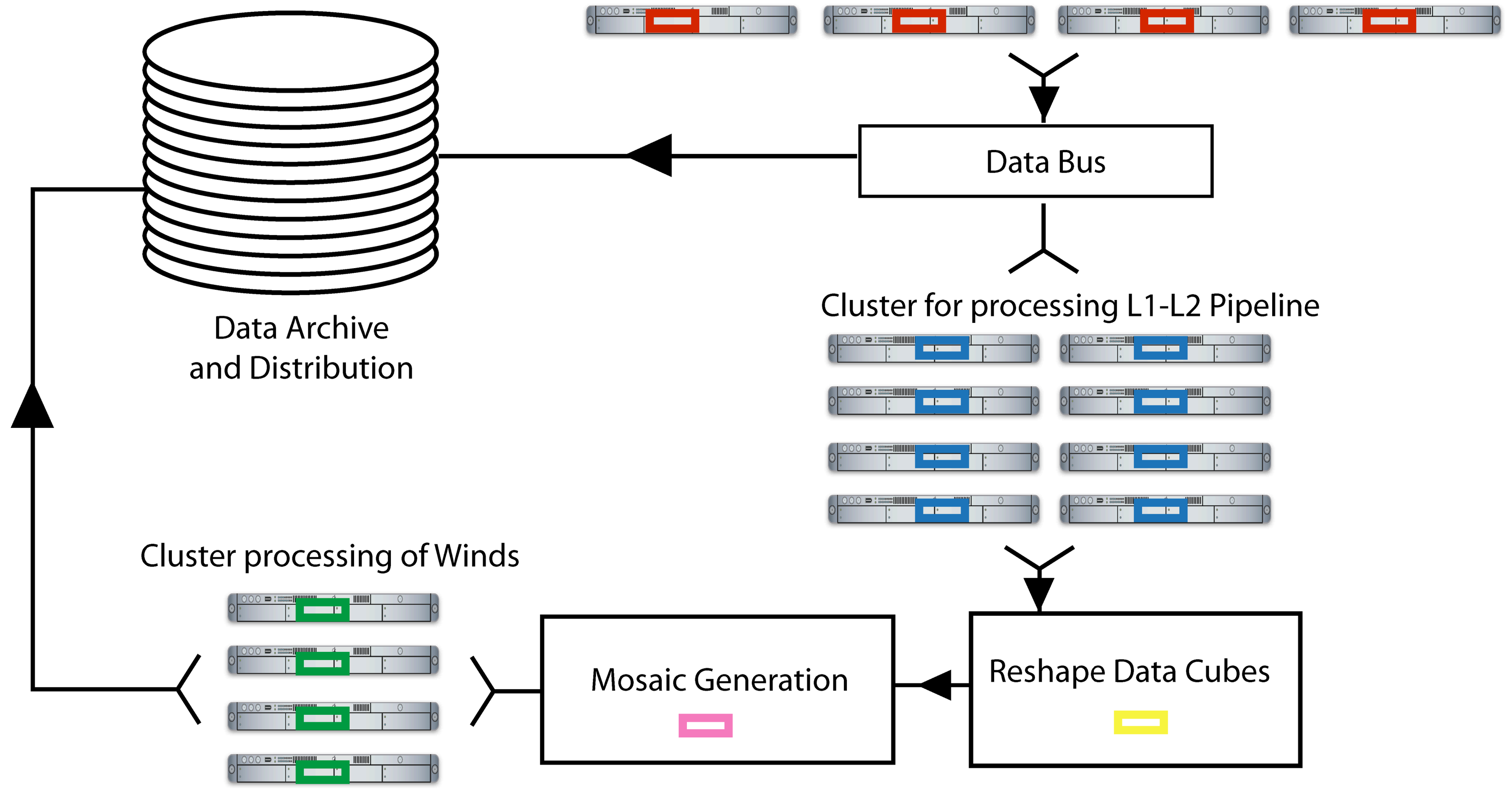


System Component Diagram

Our Approach:
- Distributed computing environment
  - Beowulf Cluster
  - High-speed interconnect
- Modern software engineering techniques
  - Component oriented design
  - Unified Modeling Language Real-Time Extension (UML-RT) used to communicate architecture and design decisions.

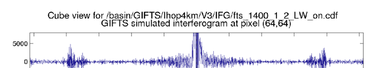
Distributing the computation:
- Data from individual pixels can be processed separately
- Dispatch data for group of pixels to worker node in cluster
- Pipelines will operate in paralllel on many pixels at once
- Audit data (monitoring channel) keeps track of
  - Algorithm versions used in processing (per-cube)
  - Node/CPU used to process pixel (per-pixel)



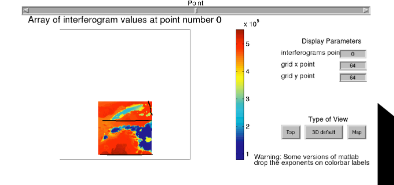


Input:

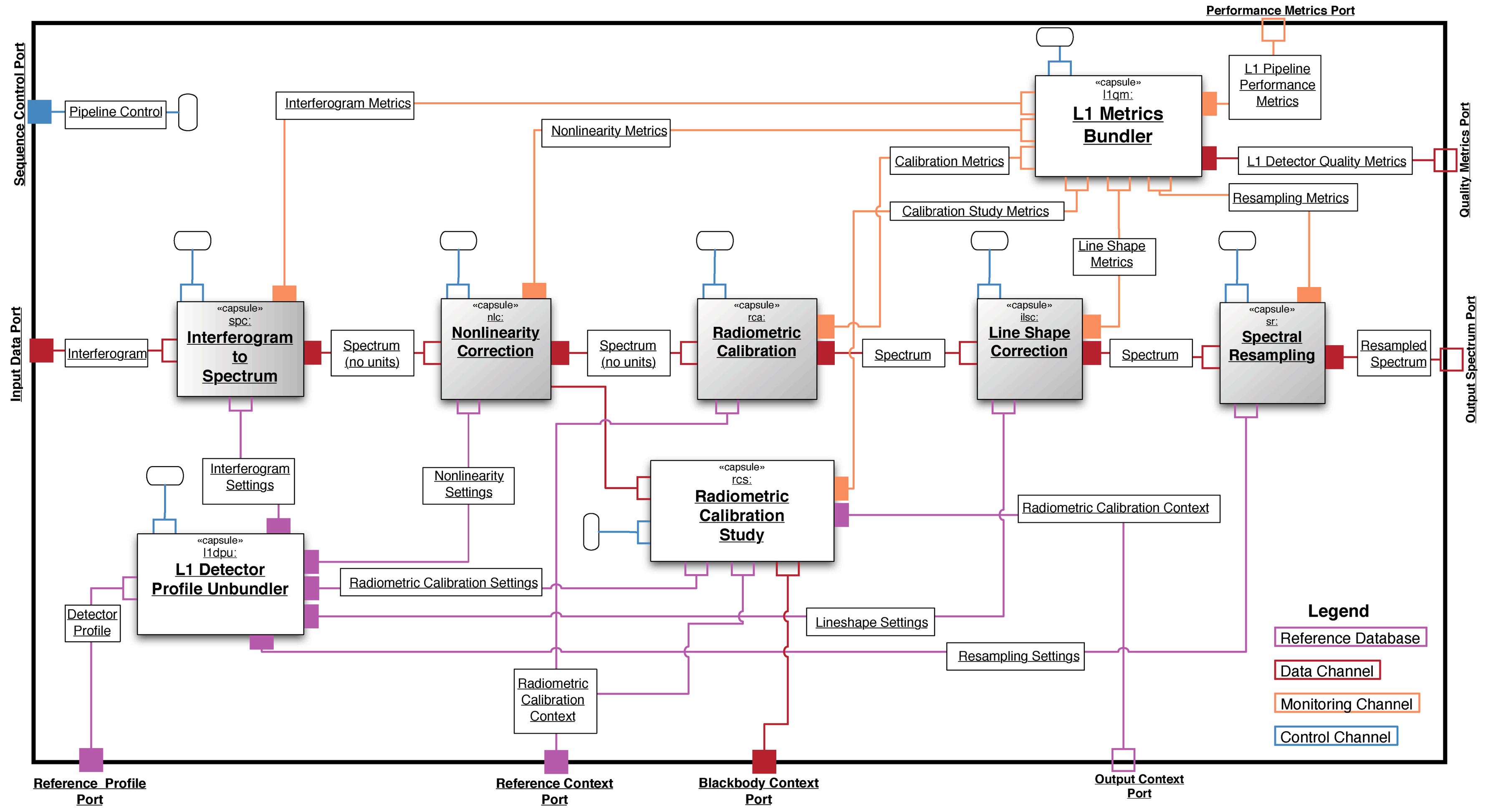
Raw Interferograms



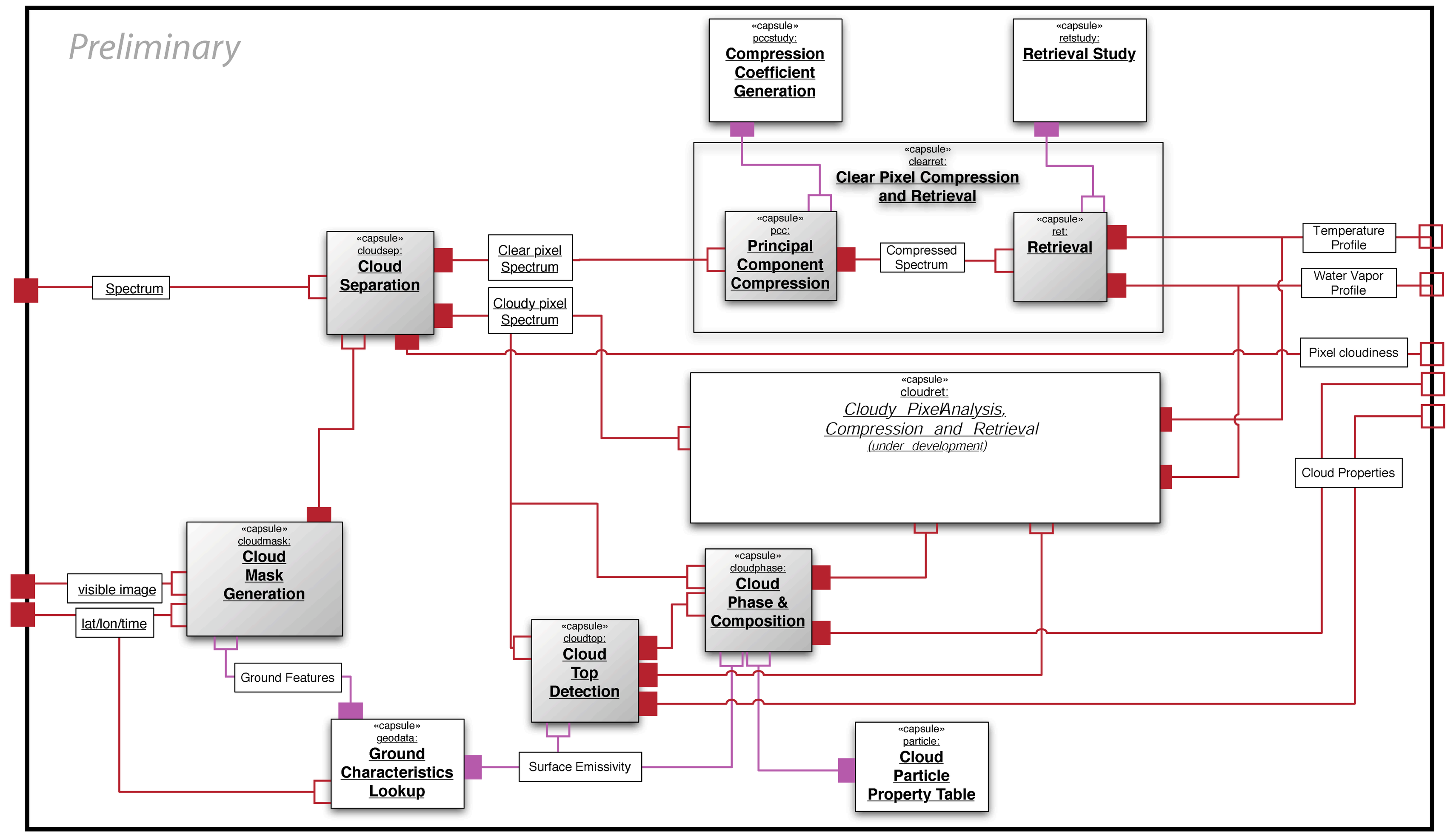
Earth Scenes



Blackbody Views (for calibration)

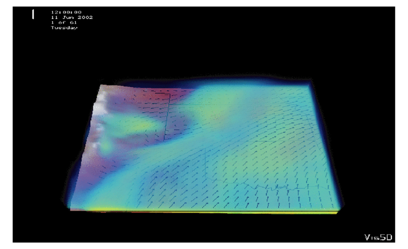


Detailed UML-RT Component Diagram of L0-L1 Processing Pipeline (per pixel)



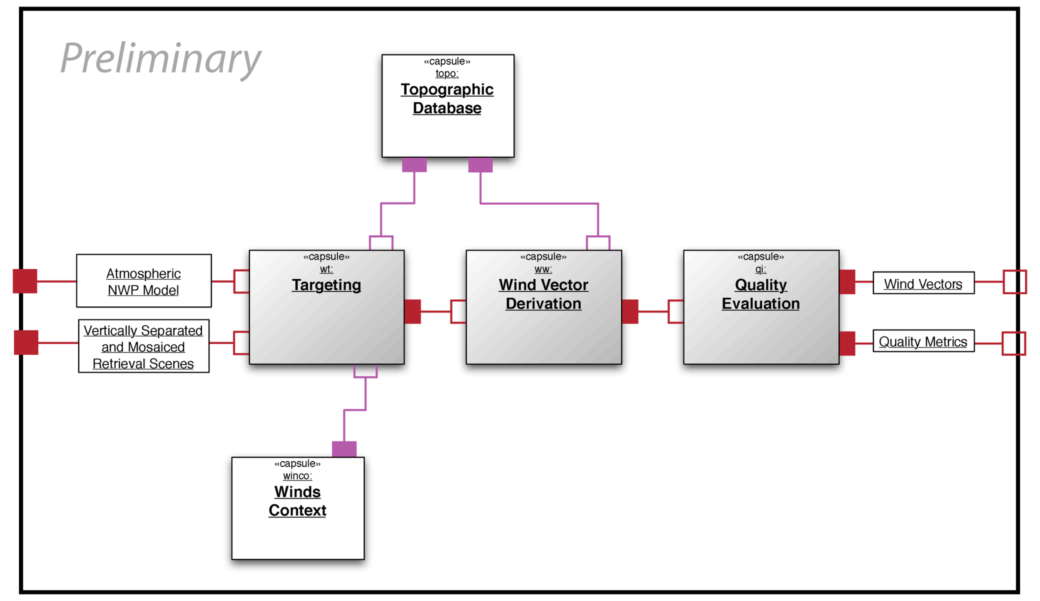
UML-RT Component Diagram of L1-L2 Processing Pipeline

Product:

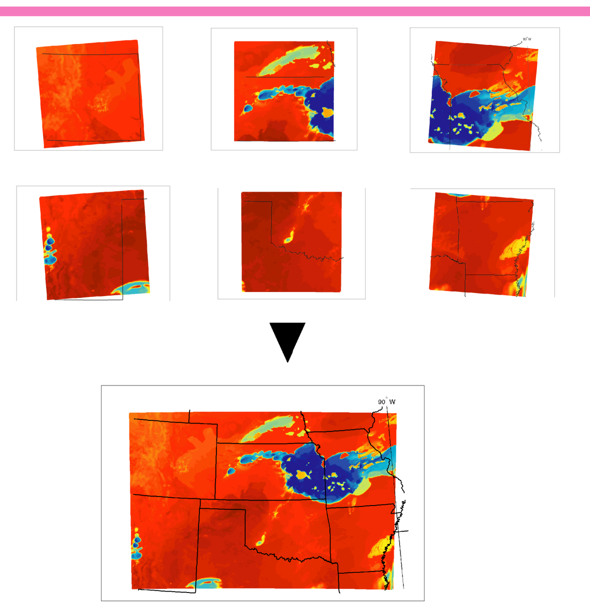


Height-separated Wind Vector Fields

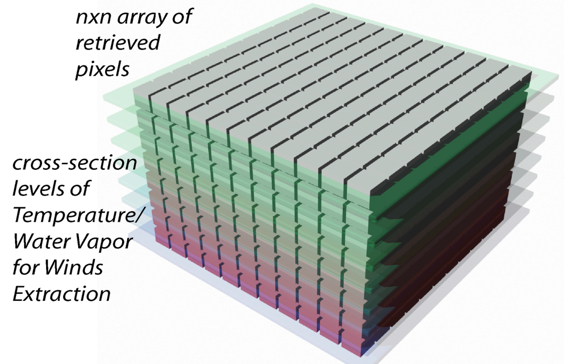
Some design tradeoffs:
- Programming framework technology MPI vs. Java/RMI vs. C++/ACE
- Performance (soft real-time) vs. robustness
- Control channel-driven (push) vs. worker node request-driven (pull) workflow.



UML-RT Component Diagram of Winds Pipeline



Mosaic Generation



- from NxN array of per-pixel T/WV retrievals
- into horizontal T/WV maps
- Needed for winds extraction

Reshape Data Cubes

Software system constraints: A new spectral resampling algorithm

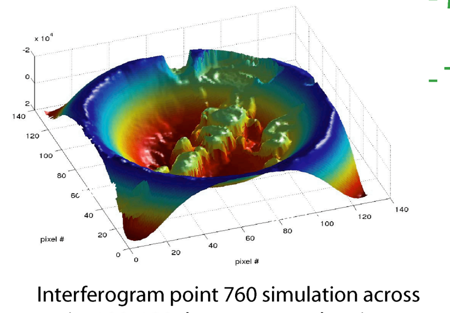
The Problem:
- Due to interferometer optics, uncorrected wavenumber scales vary for different pixels in concentric rings.

Traditional approach: double FFT to obtain upsampled spectrum, linear interpolate to common wavenumber scale

Alternate approach: Multiply spectrum against a precomputed sinc convolution matrix.

Advantages of alternate approach:
- Multiplication by precomputed matrix is > 3x faster on modern processors.
- Traditional approach was most computationally costly step in L0-L1 processing.

Drawbacks / design challenges:
- Generation of convolution matrices is much slower, necessitating caching strategy.
- Matrices are large (40 MB each), thus cannot all be stored on each worker node.



Interferogram point 760 simulation across entire 128x128 detector array showing dependence on distance from center

Solution: Pixel affinity
- correction determined by distance of pixel from center of detector array
- cache a subset of matrices on each node
- send pixels to nodes which already have the required matrices cached
- may require greater degree of centralized pixel dispatch logic

We acknowledge the support of NOAA NESDIS in this project.