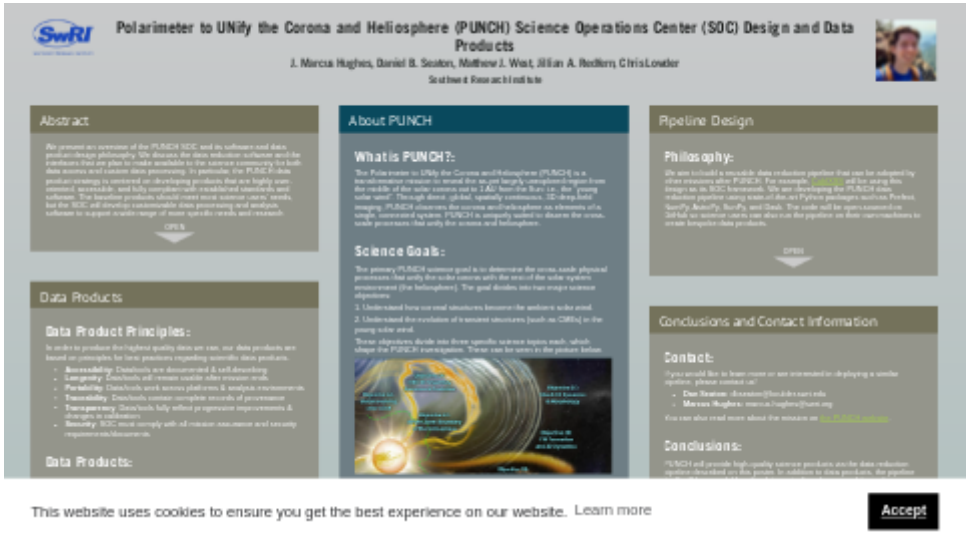
Polarimeter to UNify the Corona and Heliosphere (PUNCH) Science Operations Center (SOC) Design and Data Products



J. Marcus Hughes, Daniel B. Seaton, Matthew J. West, Jillian A. Redfern, Chris Lowder

Southwest Research Institute

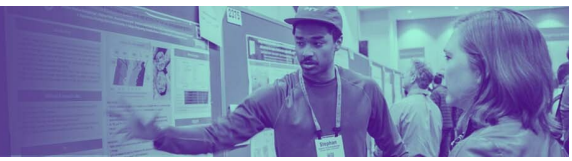




PRESENTED AT:

AGU FALL MEETING
New Orleans, LA & Online Everywhere
13–17 December 2021

Poster Gallery
brought to you by
WILEY



ABSTRACT

We present an overview of the PUNCH SOC and its software and data product design philosophy. We discuss the data reduction software and the interfaces that we plan to make available to the science community for both data access and custom data processing. In particular, the PUNCH data product strategy is centered on developing products that are highly user-oriented, accessible, and fully compliant with established standards and software. The baseline products should meet most science users' needs, but the SOC will develop customizable data processing and analysis software to support a wide range of more specific needs and research activities as well. These tools will be developed in Python using packages such as SunPy, NumPy, Dask, and Prefect. We discuss opportunities to extend our data pipeline and processing framework to other instruments and missions, as well as the possibility of generating cross-mission hybrid data products.

DATA PRODUCTS

Data Product Principles:

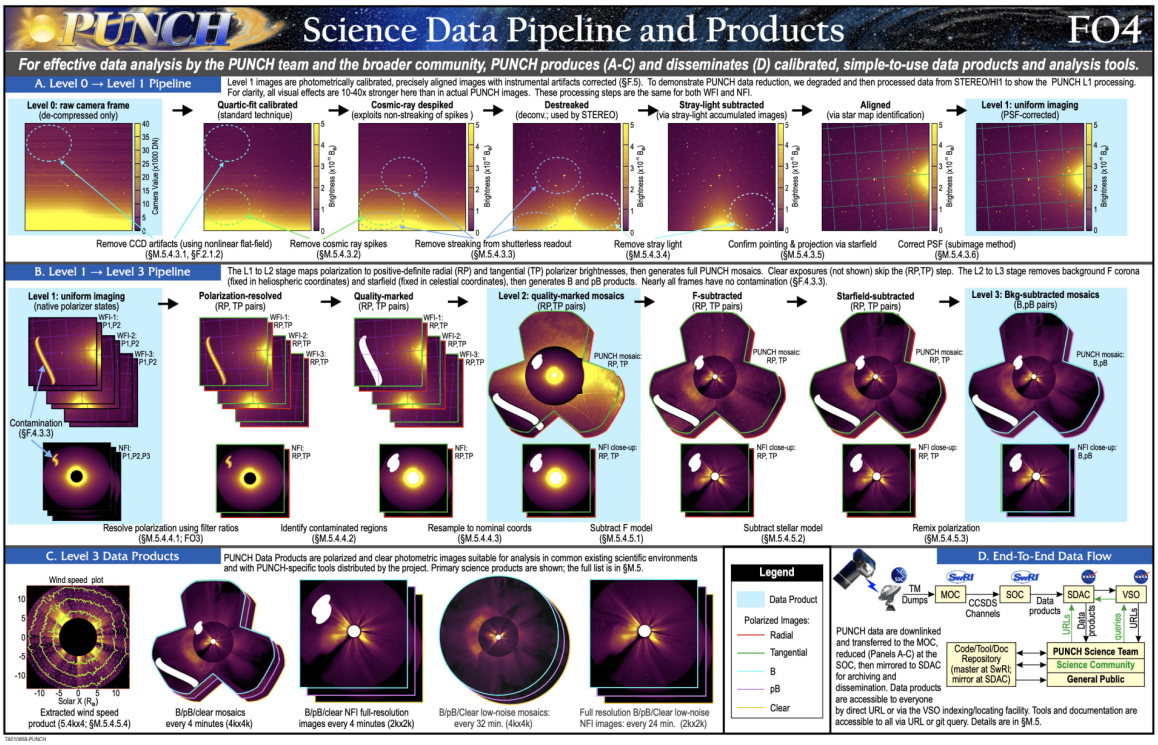
In order to produce the highest quality data we can, our data products are based on principles for best practices regarding scientific data products.

- **Accessibility**: Data/tools are documented & self-describing
- **Longevity**: Data/tools will remain usable after mission ends
- **Portability**: Data/tools work across platforms & analysis environments
- **Traceability**: Data/tools contain complete records of provenance
- **Transparency**: Data/tools fully reflect progressive improvements & changes in calibration
- **Security**: SOC must comply with all mission assurance and security requirements/documents

Data Products:

Science products are produced at several levels: 0, 1, 2, and 3.

- Level 0 products are raw camera frames. The only processing that has been applied is converting them from the CCSDS format into FITS.
- Level 1 images have been calibrated, cosmic-ray despiked, destreaked, stray-light subtracted, and aligned.
- Level 2 images are mosaics resolved into standard polarization states with bad quality pixels flagged. We produce these images both as a mosaic containing all of the observations and as only NFI close-up images.
- Level 3 products are derived and include: extracted wind speed product, mosaics every 4 minutes, NFI full-resolution every 4 minutes. low-noise mosaics every 32 minutes, and low-noise NFI images every 24 minutes.



ABOUT PUNCH

What is PUNCH?:

The Polarimeter to UNify the Corona and Heliosphere (PUNCH) is a transformative mission to reveal the as-yet largely unexplored region from the middle of the solar corona out to 1 AU from the Sun: i.e., the “young solar wind”. Through direct, global, spatially continuous, 3D deep-field imaging, PUNCH observes the corona and heliosphere as elements of a single, connected system. PUNCH is uniquely suited to discern the cross-scale processes that unify the corona and heliosphere.

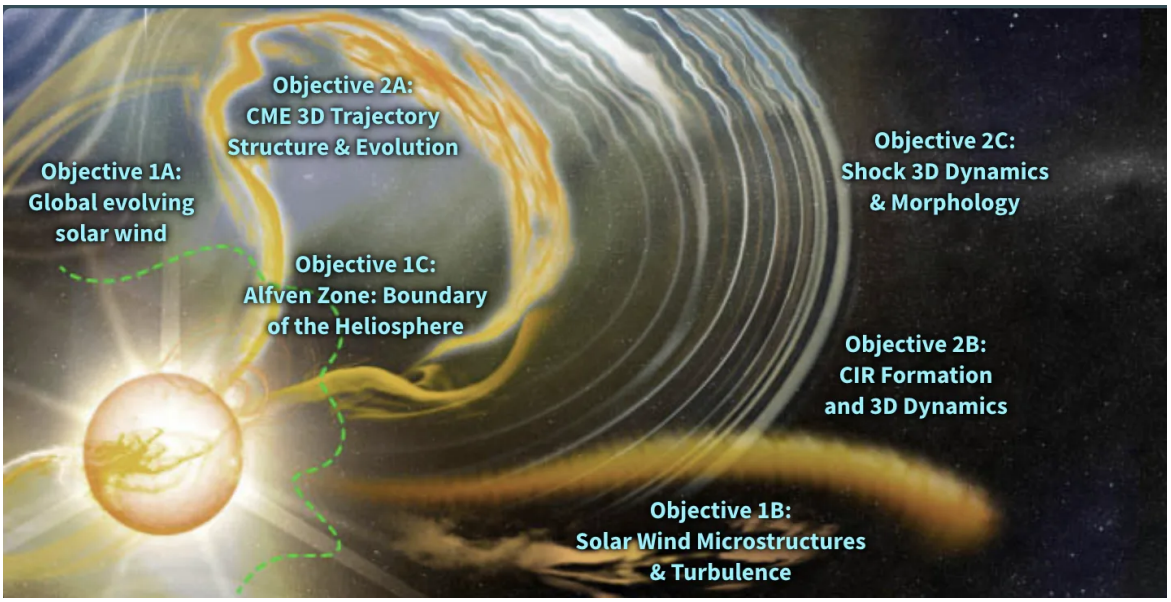
Science Goals:

The primary PUNCH science goal is to determine the cross-scale physical processes that unify the solar corona with the rest of the solar system environment (the heliosphere). The goal divides into two major science objectives:

1. Understand how coronal structures become the ambient solar wind.

2. Understand the evolution of transient structures (such as CMEs) in the young solar wind.

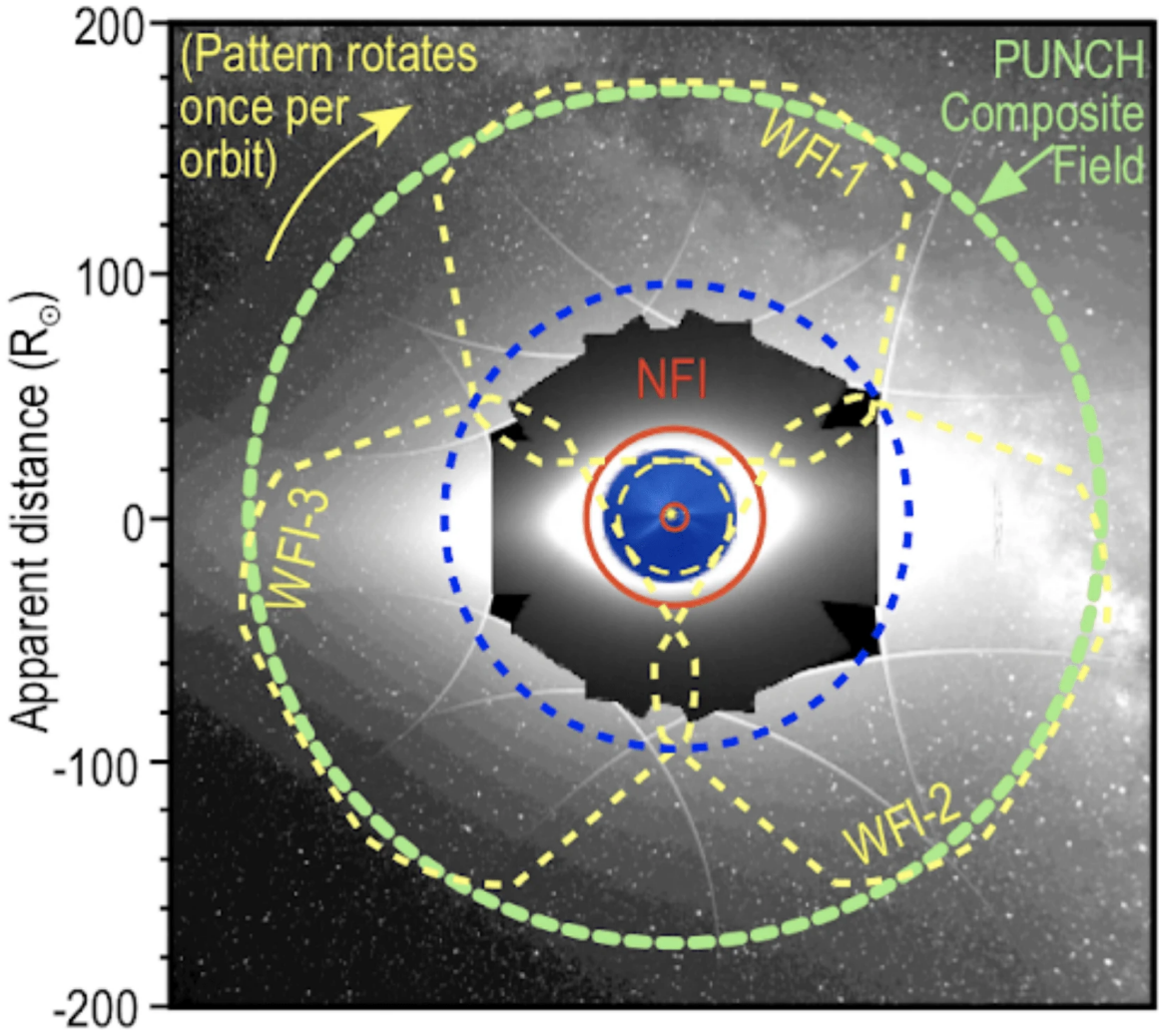
These objectives divide into three specific science topics each, which shape the PUNCH investigation. These can be seen in the picture below.



More detail about science goals can be found on the PUNCH website (https://punch.space.swri.edu/).

Instrumentation:

Four satellites are necessary because PUNCH views the entire inner solar system continuously, and Earth blocks approximately half the view from any one location in orbit. The lone Narrow Field Imager (NFI, shown in red in the figure below) acquires an annular field of view (FOV) around the Sun. Three Wide Field Imagers (WFIs, shown in yellow in the figure below) acquire data in a trefoil on the sky. As the spacecraft orbit, the trefoil rotates on the sky and builds up the full 90° circular PUNCH FOV.



Through the data reduction pipeline, these are constrcuted into one composite field (the green circle in the figure above).

PIPELINE DESIGN

Philosophy:

We aim to build a reusable data reduction pipeline that can be adopted by other missions after PUNCH. For example, CubIXSS (https://www.swri.org/press-release/nasa-selects-swriled-cubesat-assess-origins-hot-plasma-suns-corona) will be using this design as its SOC framework. We are developing the PUNCH data reduction pipeline using state-of-the-art Python packages such as Prefect, NumPy, AstroPy, SunPy, and Dask. The code will be open-sourced on GitHub so science users can also run the pipeline on their own machines to create bespoke data products.

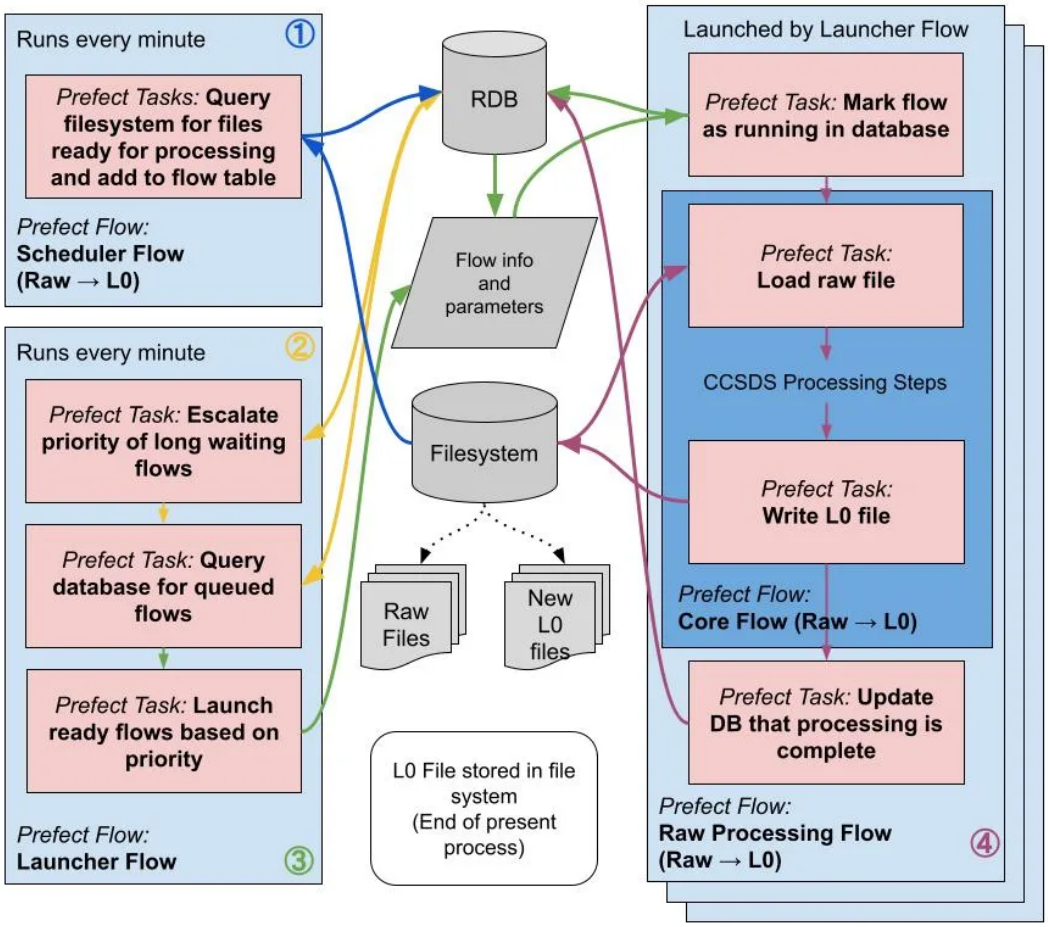
Using Prefect:

Prefect (https://www.prefect.io/) is a dataflow orchestrator that handles pipeline scheduling, concurrency, logging, error handling, and many other tasks. At its core, it reduces the pipeline design into a set of *tasks* organized into a *flow*. A task is a basic unit of computation, e.g. loading an image, despiking an image, subtracting darks. These tasks are connected together into a Prefect flow so they can run in a defined order. We envision at least two flow to execute each processing level of the pipeline.

Prefect Flows:

Our design consists of several types of flows that we create:

1. **Core flows:** Science processing code ran to produce a data product.
2. **Processing flows:** A wrapper around the core flow that also executes tasks specific to our automated pipeline, e.g. updating databases.
3. **Scheduler flows**: A kind of flow that will schedule processing flows to run as raw data arrives. This is done using a database of scheduled flows that get kicked off by the launcher flow.
4. **A launcher flow:** The controller that initiates launcher flows as compupting resources become available.



Above is a representative overview of Prefect implementation of the Raw→L0 processing segment, including the scheduler flow (1, blue arrows; top left), launcher flow (2, yellow arrows & 3, green arrows; bottom left), and processing flow (4, purple arrows; right) as well as interfaces between the pipeline elements and to the filesystem and other auxiliary services in between. Processing begins at 1 with the scheduler flows and progresses to 4 with the processing flow, passing data through the different colored lines. Each different color for the lines and corresponding label numbers indicates a different stage in the processing pipeline. The dotted lines indicate connections between elements of the pipeline. This model of computation is repeated for each processing level.

CONCLUSIONS AND CONTACT INFORMATION

Contact:

If you would like to learn more or are interested in deploying a similar pipeline, please contact us!

- **Dan Seaton:** dbseaton@boulder.swri.edu
- **Marcus Hughes:** marcus.hughes@swri.org

You can also read more about the mission on the PUNCH website (https://punch.space.swri.edu/index.php).

Conclusions:

PUNCH will provide high-quality science products via the data reduction pipeline described on this poster. In addition to data products, the pipeline itself will be a model for other future pipelines because of its modern design.

AUTHOR INFORMATION

Team from Southwest Research Institute Boulder:

J. Marcus Hughes

Daniel B. Seaton

Matthew J. West

Jillian A. Redfern

Chris Lowder

ABSTRACT

We present an overview of the PUNCH SOC and its software and data product design philosophy. We discuss the data reduction software and the interfaces that we plan to make available to the science community for both data access and custom data processing. In particular, the PUNCH data product strategy is centered on developing products that are highly user-oriented, accessible, and fully compliant with established standards and software. The baseline products should meet most science users' needs, but the SOC will develop customizable data processing and analysis software to support a wide range of more specific needs and research activities as well. These tools will be developed in Python using packages such as SunPy, NumPy, Dask, and Prefect. We discuss opportunities to extend our data pipeline and processing framework to other instruments and missions, as well as the possibility of generating cross-mission hybrid data products.