PUNCH Science Operations Center Data Products and Software

Marcus Hughes
PUNCH SOC Lead

DASH Meeting
14-16 Oct 2024

SOC Members

Derek Lamb

Dan Seaton

Ritesh Patel

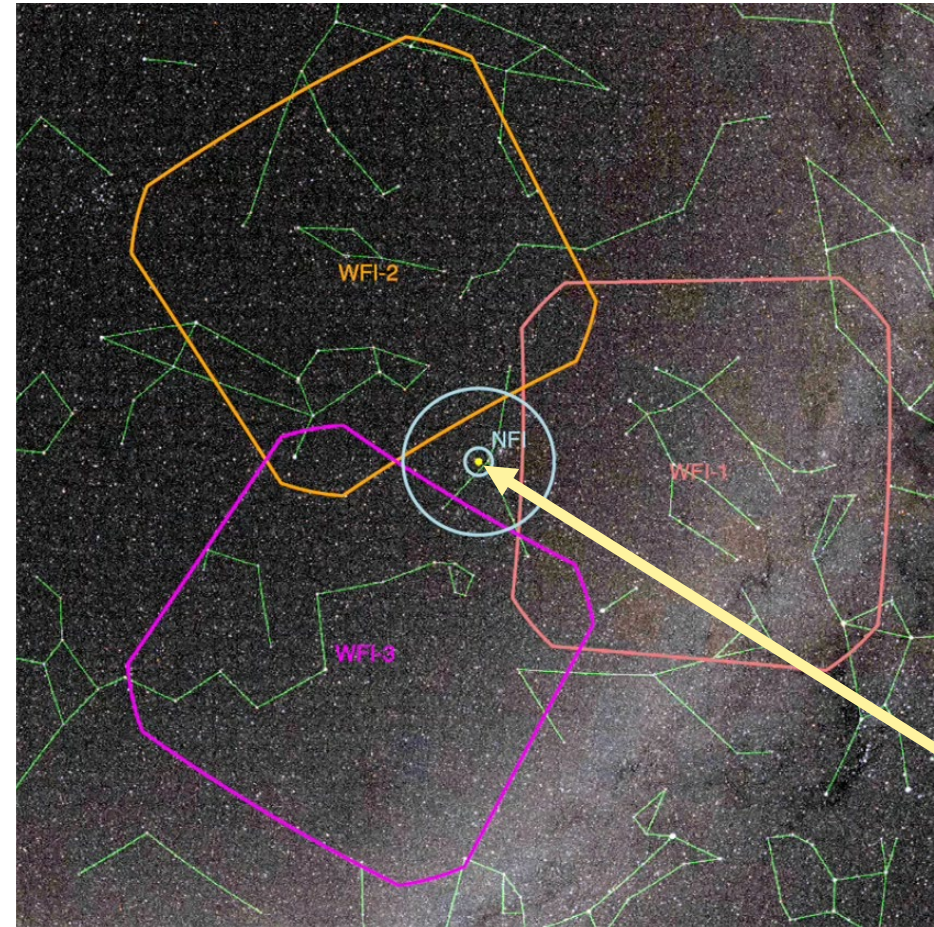Marcus Hughes

Matt West

Chris Lowder

Sarah Kovac

- PUNCH will be a constellation of four smallsats to study the solar wind

- Launch early next year

- Complicated processing pipeline of stitching together images from multiple satellites

- Software will be completely open

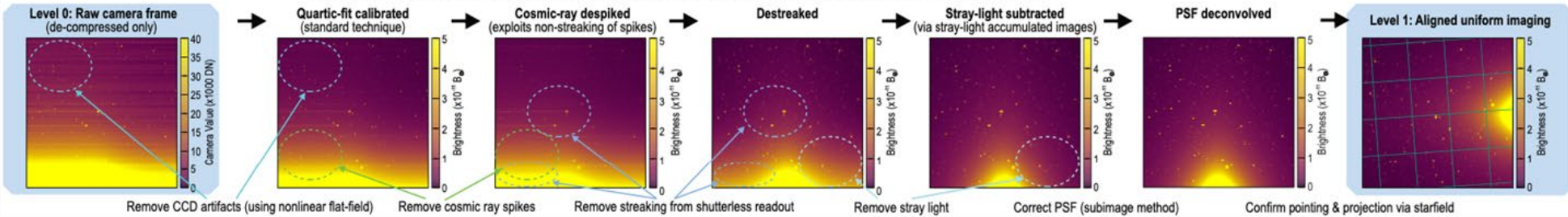PUNCH will change how we understand the solar wind

Science Data Pipeline and Products

For effective data analysis by the PUNCH team and the broader community, PUNCH produces (A-C) and disseminates (D) calibrated, simple-to-use data products and analysis tools.
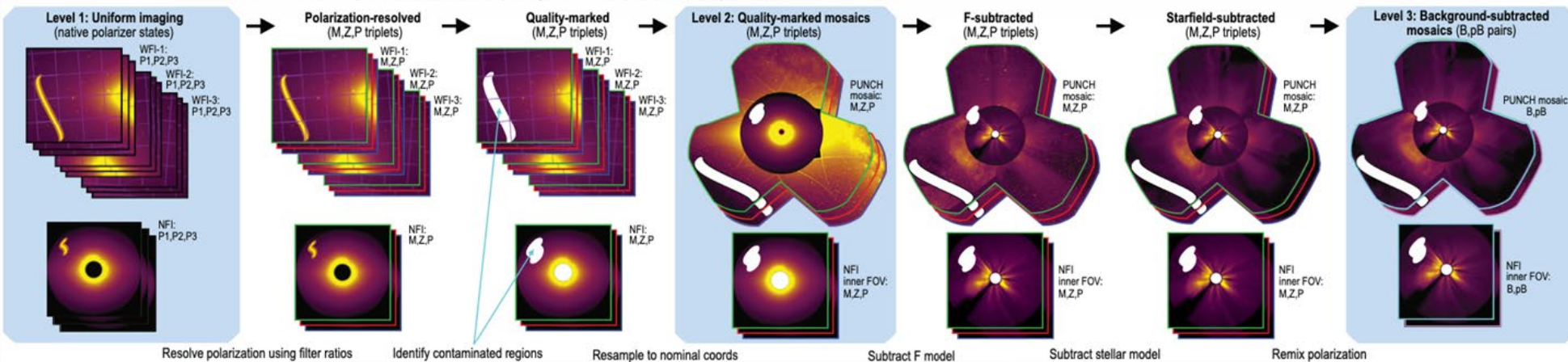
A. Level 0 → Level 1 Pipeline

Level 1 images are photometrically calibrated, precisely aligned images with instrumental artifacts corrected. To demonstrate PUNCH data reduction, we degraded and then processed data from STEREO/HI1 to show the PUNCH L1 processing. For clarity, all visual effects are 10-40x stronger here than in actual PUNCH images. These processing steps are the same for both WFI and NFI.

- Level 0: Raw camera frame (de-compressed only) — Remove CCD artifacts (using nonlinear flat-field)
- Quartic-fit calibrated (standard technique) — Remove cosmic ray spikes
- Cosmic-ray despiked (exploits non-streaking of spikes) — Remove streaking from shutterless readout
- Destreaked — Remove stray light
- Stray-light subtracted (via stray-light accumulated images) — Correct PSF (subimage method)
- PSF deconvolved
- Level 1: Aligned uniform imaging — Confirm pointing & projection via starfield
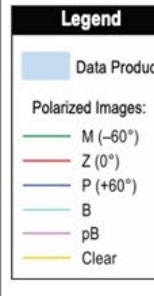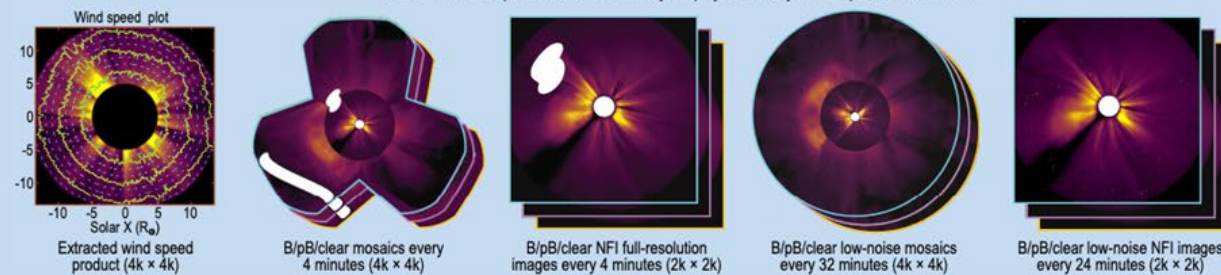
B. Level 1 → Level 3 Pipeline

The L1 to L2 stage maps polarization to M,Z,P triplet polarizer brightnesses, then generates full PUNCH mosaics. Clear exposures (not shown) skip the (M,Z,P) step. The L2 to L3 stage removes background F corona (fixed in heliospheric coordinates) and starfield (fixed in celestial coordinates), then generates B and pB products. Nearly all frames have no contamination.

- Level 1: Uniform imaging (native polarizer states) — Resolve polarization using filter ratios
- Polarization-resolved (M,Z,P triplets) — Identify contaminated regions
- Quality-marked (M,Z,P triplets) — Resample to nominal coords
- Level 2: Quality-marked mosaics (M,Z,P triplets) — Subtract F model
- F-subtracted (M,Z,P triplets) — Subtract stellar model
- Starfield-subtracted (M,Z,P triplets) — Remix polarization
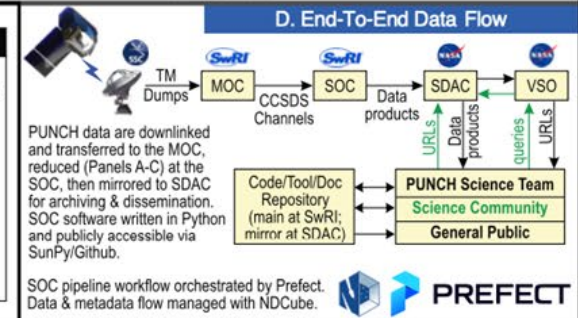- Level 3: Background-subtracted mosaics (B,pB pairs)

C. Level 3 Data Products

PUNCH Data Products are polarized and clear photometric images suitable for analysis in common existing scientific environments and with PUNCH-specific tools distributed by the project. Primary science products are shown.

- Extracted wind speed product (4k × 4k)
- B/pB/clear mosaics every 4 minutes (4k × 4k)
- B/pB/clear NFI full-resolution images every 4 minutes (2k × 2k)
- B/pB/clear low-noise mosaics every 32 minutes (4k × 4k)
- B/pB/clear low-noise NFI images every 24 minutes (2k × 2k)

Legend

Data Product

Polarized Images:
- M (−60°)
- Z (0°)
- P (+60°)
- B
- pB
- Clear

D. End-To-End Data Flow

PUNCH data are downlinked and transferred to the MOC, reduced (Panels A-C) at the SOC, then mirrored to SDAC for archiving & dissemination. SOC software written in Python and publicly accessible via SunPy/Github.

SOC pipeline workflow orchestrated by Prefect. Data & metadata flow managed with NDCube.

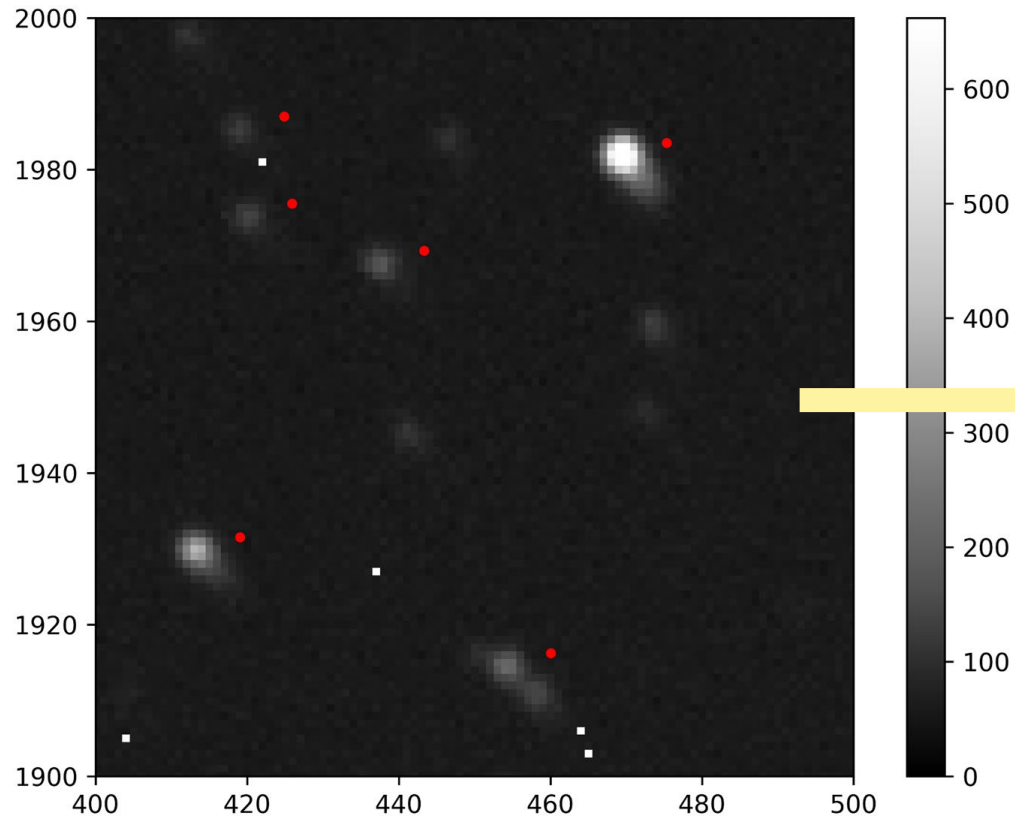Original figures and layout from PUNCH CSR

What Makes PUNCH Software Special?

- Implements science code completely separate from automation logic
  - Enables more scientist interaction and easily added custom flows
  - Can be run on any machine with or without database set up
- Tracks uncertainty
  - Every step of the pipeline tracks the uncertainty in observed values so it can be reported with the observations
- Allows for reprocessing of data while prioritizing space weather products
- Includes novel algorithms
  - Point spread function regularization
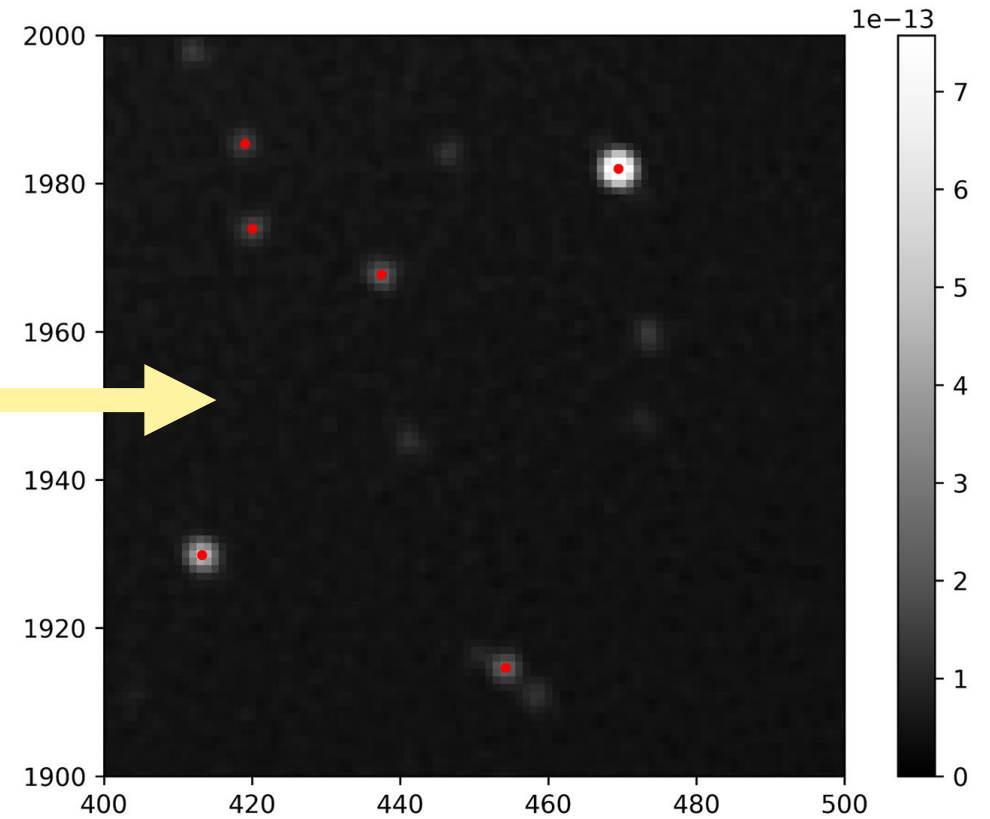 - WCS alignment
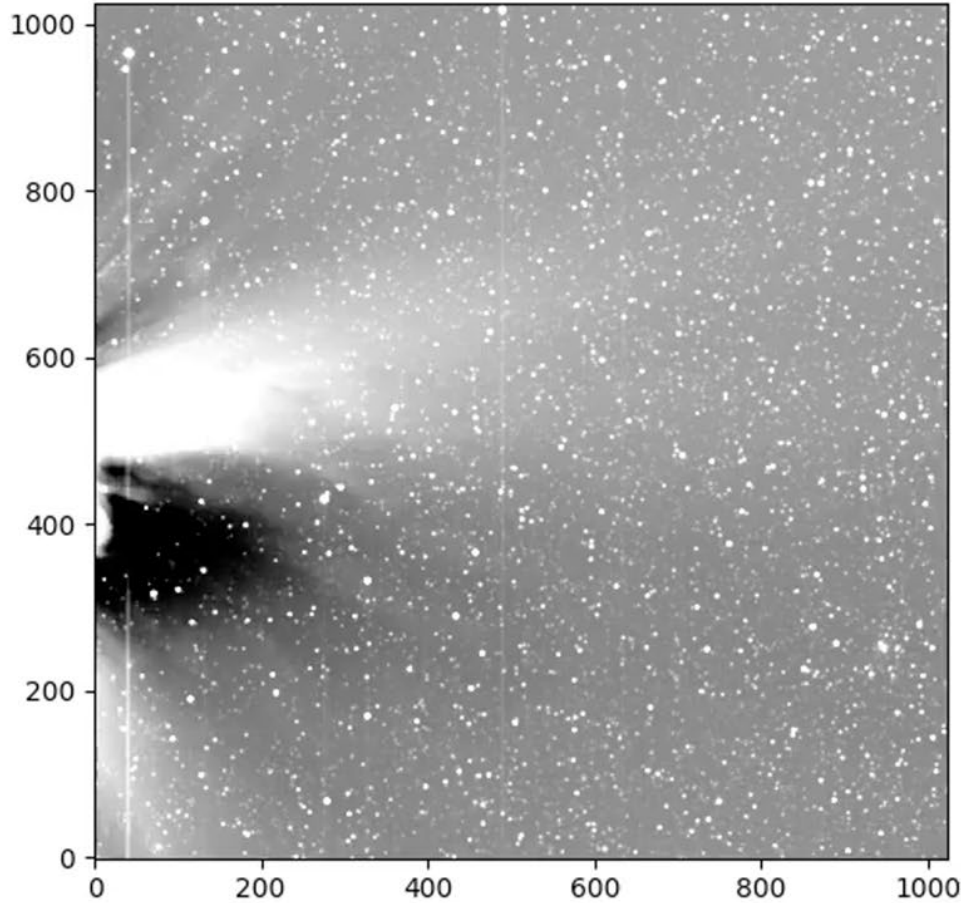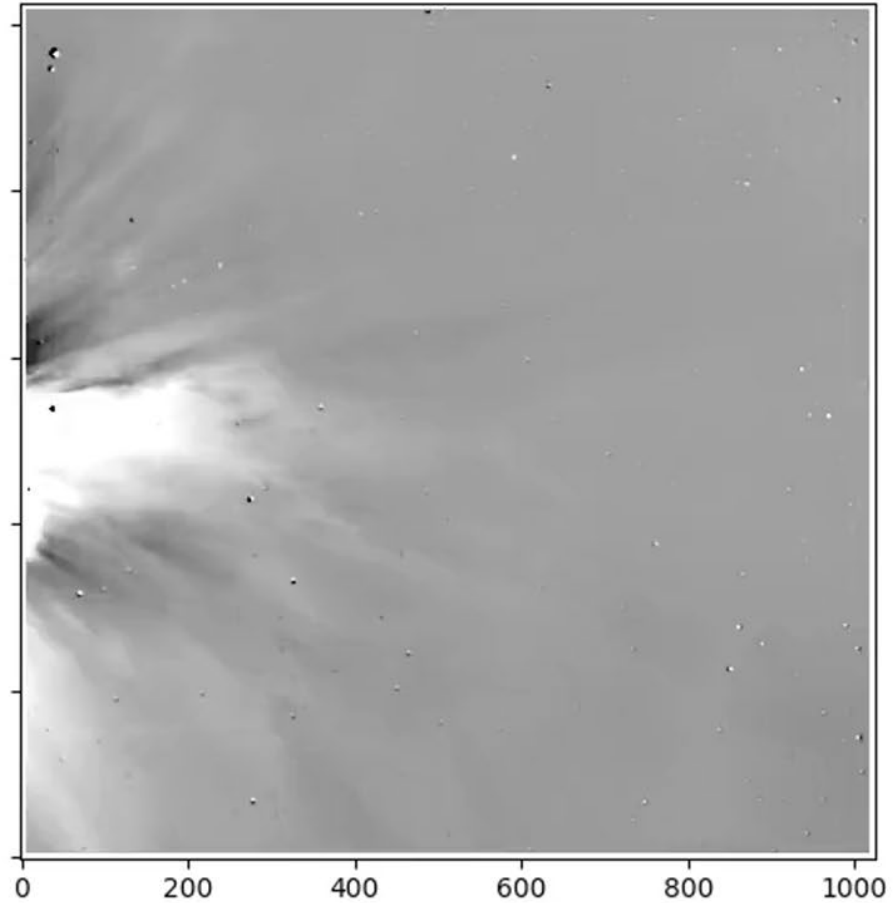 - Starfield removal

Level 0

Level 1

Sneak peek: Removing Stars

Before
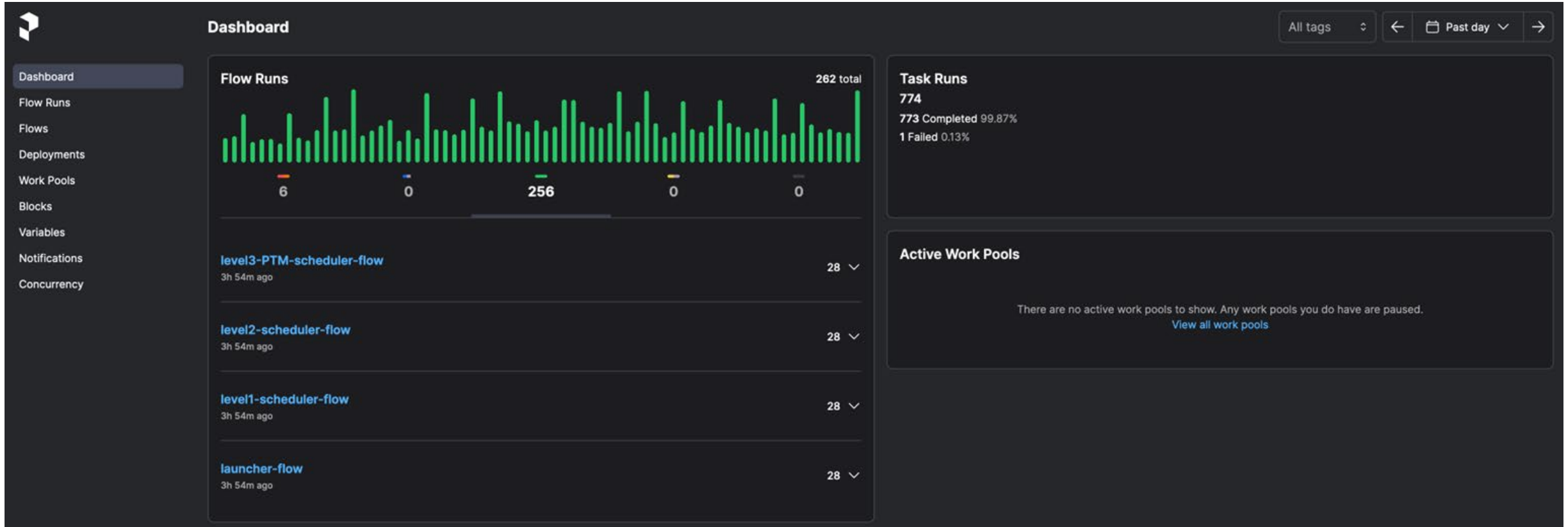
After

Prefect: the Orchestrator behind the PUNCH pipeline

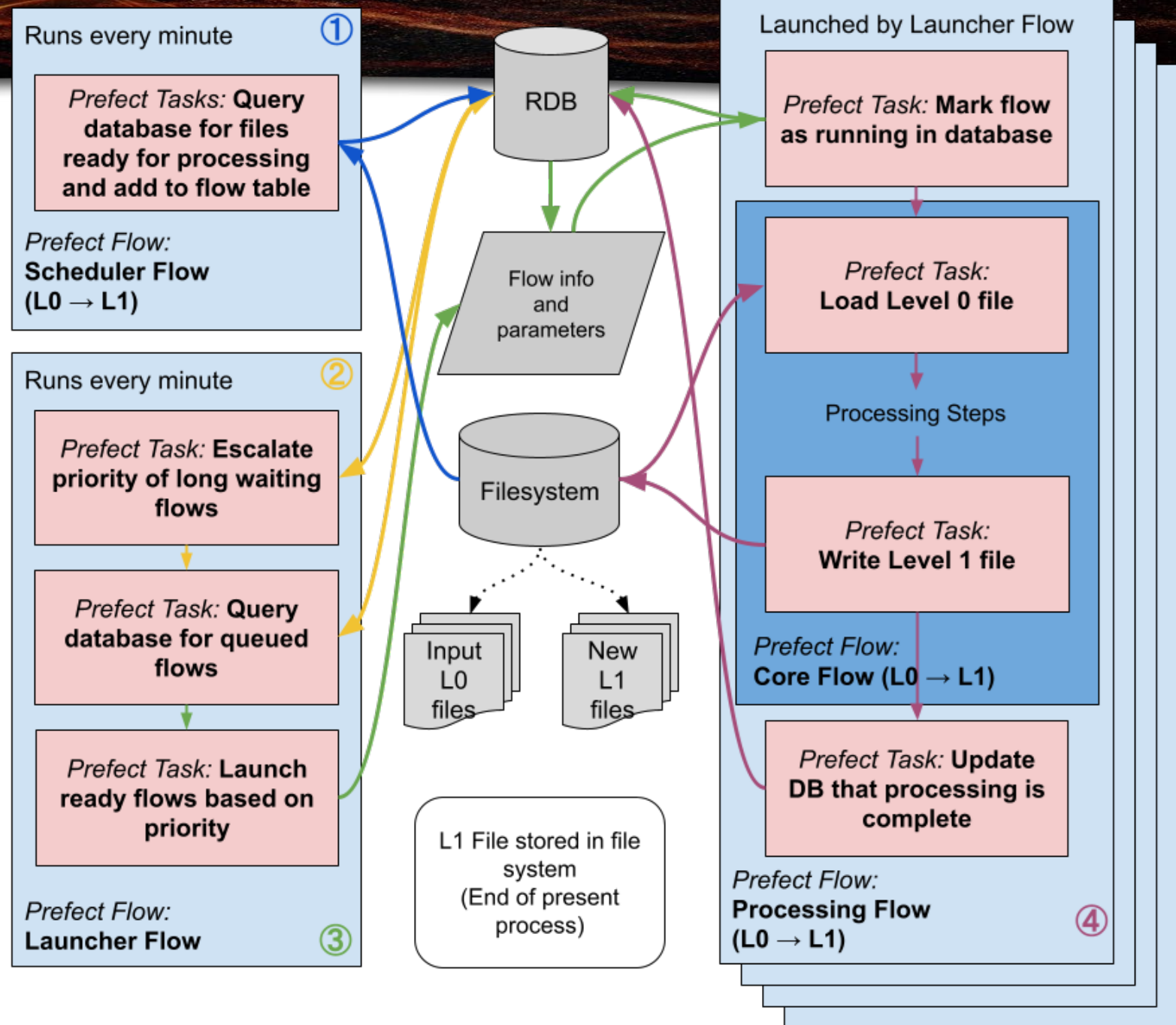Prefect provides easy observability and automation.

All science code exists in a package called *punchbowl* and is represented by the dark blue

Automation is in *punchpipe* package

4 types of flows:
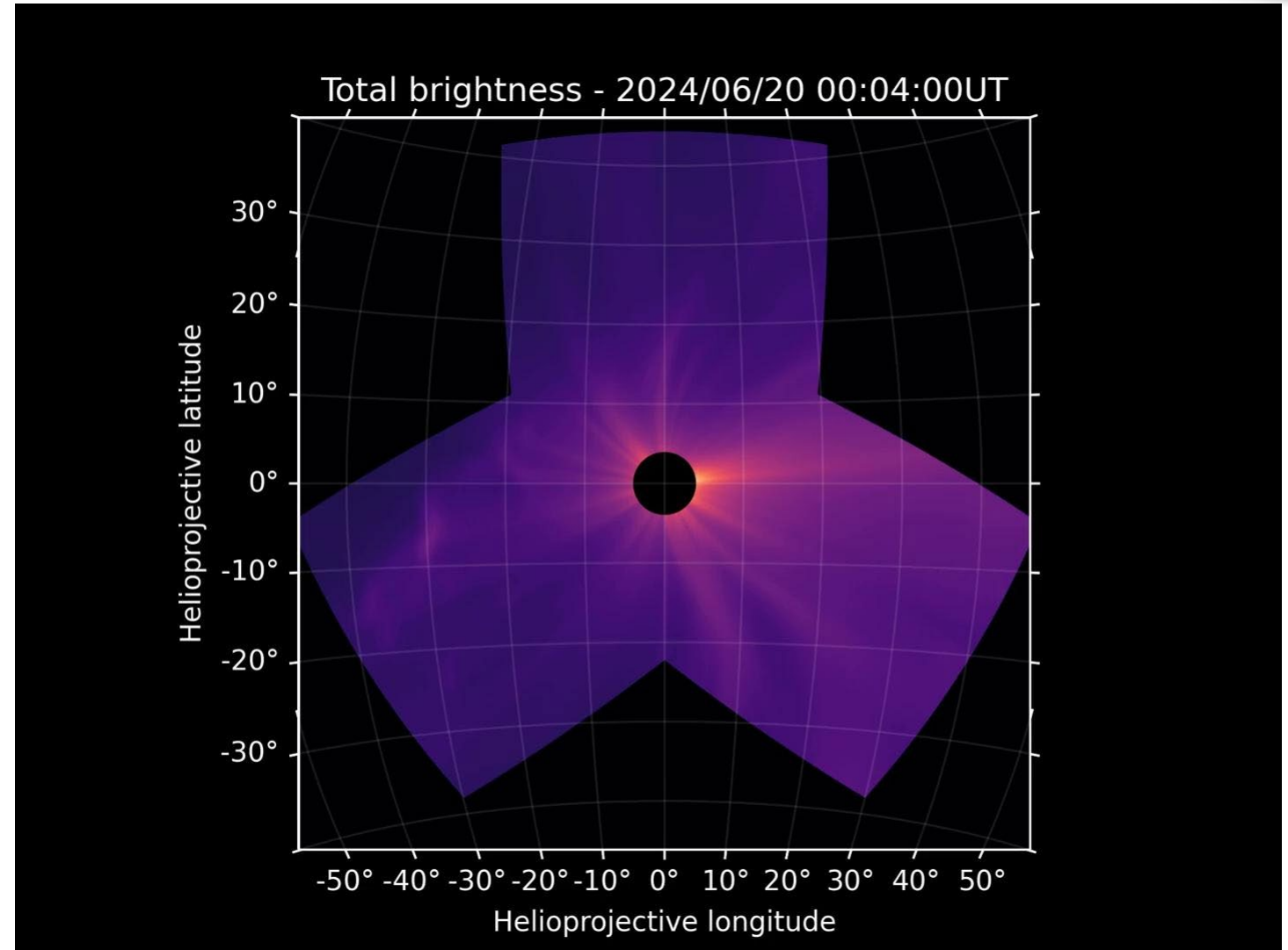1. Core
2. Processing
3. Scheduler
4. Launcher

① Scheduler Flow (L0 → L1) — Runs every minute
- *Prefect Tasks:* **Query database for files ready for processing and add to flow table**

② Launcher Flow — Runs every minute
- *Prefect Task:* **Escalate priority of long waiting flows**
- *Prefect Task:* **Query database for queued flows**
- *Prefect Task:* **Launch ready flows based on priority**

③ Launcher Flow

RDB

Flow info and parameters

Filesystem
- Input L0 files
- New L1 files

L1 File stored in file system (End of present process)

④ Processing Flow (L0 → L1) — Launched by Launcher Flow
- *Prefect Task:* **Mark flow as running in database**
- Core Flow (L0 → L1):
  - *Prefect Task:* **Load Level 0 file**
  - Processing Steps
  - *Prefect Task:* **Write Level 1 file**
- *Prefect Task:* **Update DB that processing is complete**

- Create simulated data early so you can test early
  - Start with bare bones simulation and gradually make it more sophisticated
- Don't neglect code documentation
- Don't overdo official documents
  - They can eat up the entire development schedule
- Continuous integration is wonderful
- Clearly define your releases and release early

Where We're Headed

- Launch early next year
- Data available as soon as commissioning ends

Total brightness - 2024/06/20 00:04:00UT