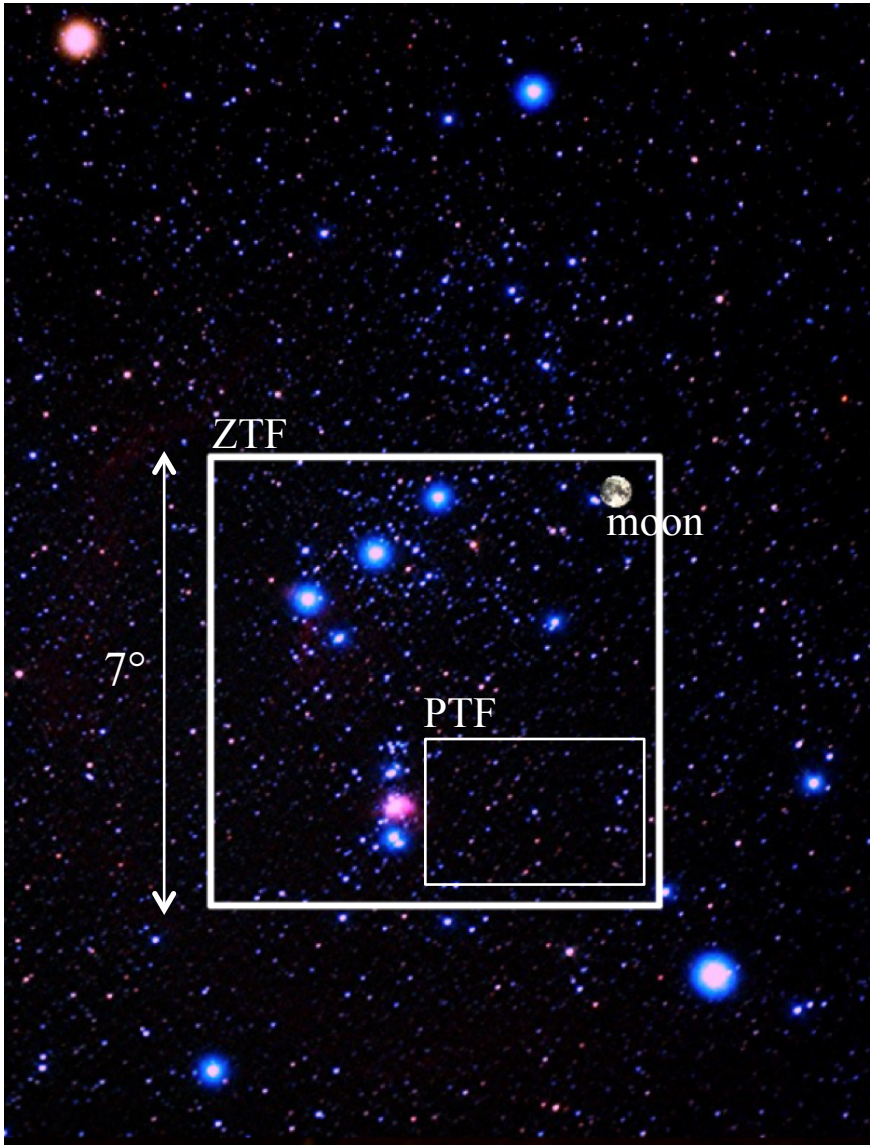
ZTF Pipelines and Deliverables

Frank Masci & the IPAC-Caltech ZTF Team

Galactic Science Workshop, March 2017

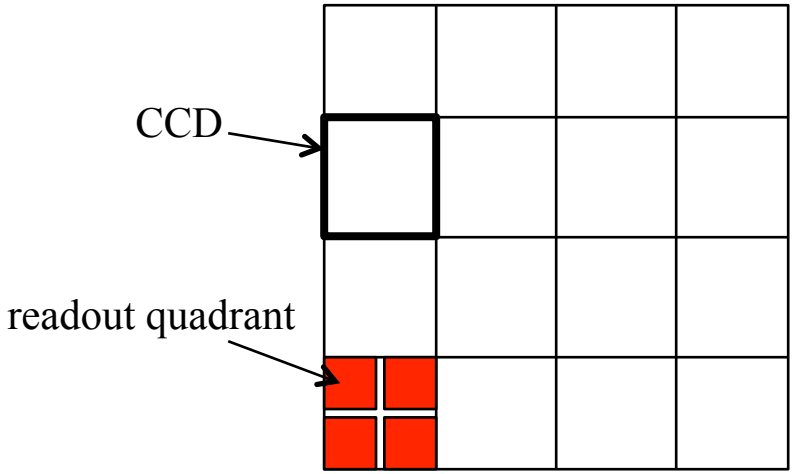
ZTF Field-of-View



- Survey rate is ~ 3760 deg^2 / hour
  Faster than the sky rotates!
- Depth: R ~ 20.4 mag AB (5σ)

ZTF Raw Camera Image Data

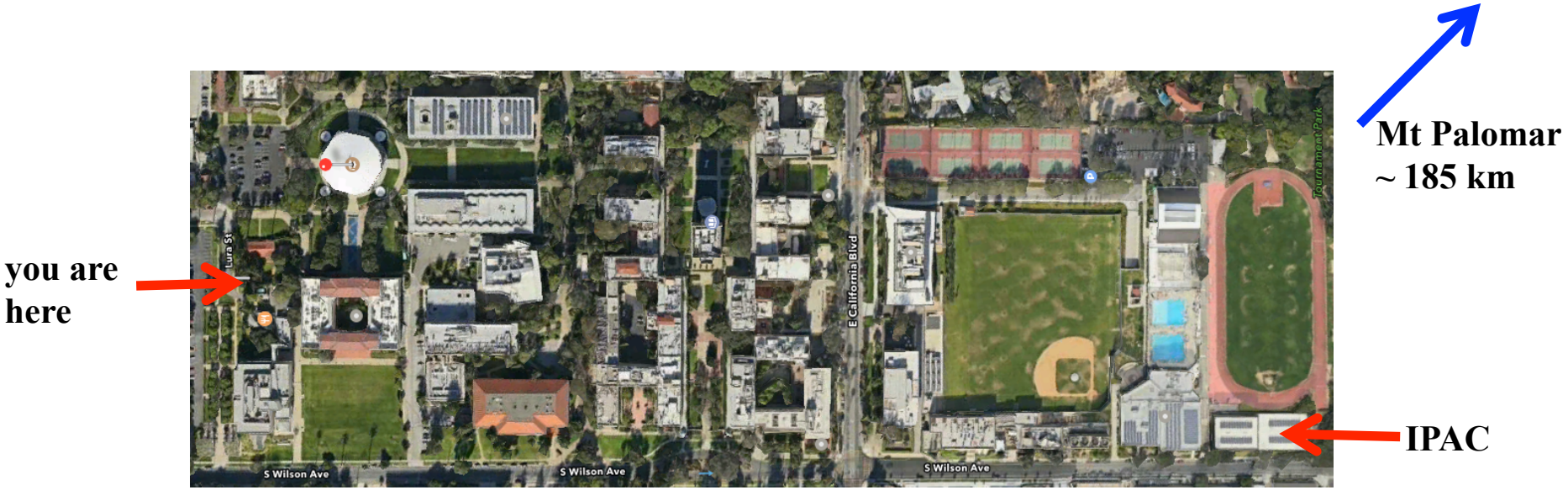
- One camera exposure: 16 CCDs; each ~ 6k x 6k pixels
- Image data packet transmitted is one CCD (four readout-quadrant images)
- 16 CCD-based image files are transmitted every 45 sec.
- Full camera exposure: ~ 1.3GB uncompressed
- Require *lossy* compression to accommodate transfer bandwidth (~ 110 - 150 Mbits/sec, variable)



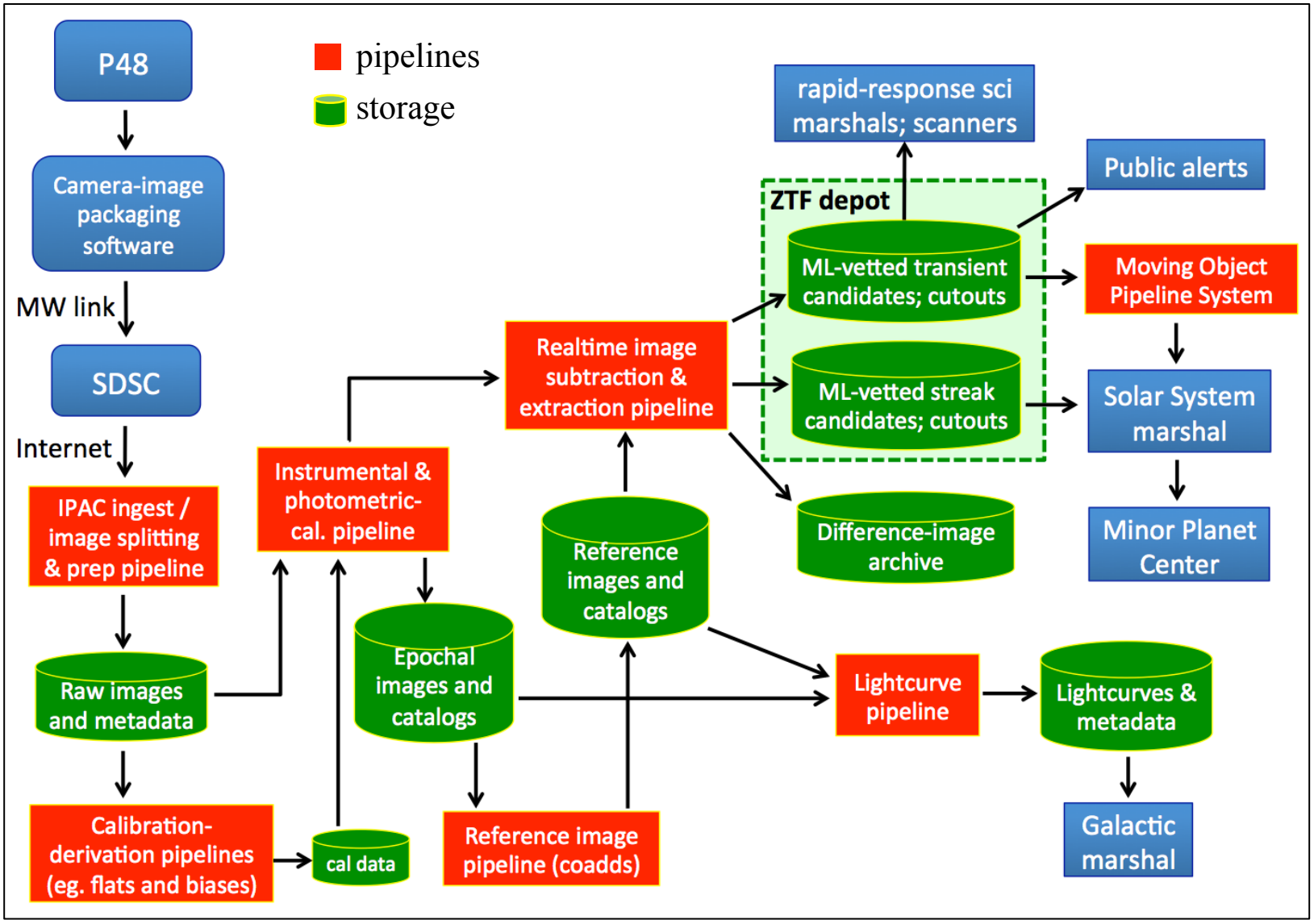
Basic image-unit for pipeline processing from which all products are derived is a ~3k x 3k readout quadrant image.

The ZTF Science Data System (ZSDS)

- The ZSDS is housed at the Infrared Processing and Analysis Center (IPAC), Caltech
- IPAC is a multi-mission science center (IRAS, ISO, *Spitzer*, WISE, Herschel, Planck, 2MASS …)
- Responsibility for ZTF (like PTF):
  - ➢ data transfer from P48 to IPAC;
  - ➢ data processing pipelines;
  - ➢ long-term data archiving, curation, user-interfaces, and APIs to retrieve data;
  - ➢ generation of transient alerts and metadata to support near real-time discovery;
  - ➢ maintenance of operations, databases, fileservers, and archive infrastructure.



Overview of the ZTF Data System



ZTF Pipelines

Overall, there are 10 inter-dependent pipelines:

Raw data ingestion/processing:

1. Raw data ingest, archival of raw images and storage of metadata in database [*realtime*]
2. Raw-image decompression, splitting into readout-quadrant images, floating bias correction, simple QA [*realtime*]

Calibration generation:

3. Bias-image derivation from stacking calibration images acquired in afternoon [*made before on-sky operations*]
4. High-ν flat (pixel-to-pixel responsivity) from stacking calibration images [*made before on-sky operations*]
5. Low-ν flat from either long-term ZPVM or dithered-star observations [*every week, month or longer?*]

Real-time:

6. Instrumental calibration of readout-quadrant images: astrometry and photmetric cal [*realtime*]
7. Image subtraction and transient discovery (point sources / streaks), metadata and cutouts [*realtime*]

Ensemble-based processing:

8. Reference-image generation (co-addition of epochal images from 6) [*as needed: when good quality data available*]
9. Source-matching with relative photometric refinement for lightcurves; inputs from 6 [*every two weeks or longer?*]
10. Moving object pipeline system (MOPS): tracklets from linking transients from 7 [*every 3 or 4 hours during night*]

Deliverables and Products

1. **Instrumentally calibrated, <u>readout-quadrant based</u> epochal image products:**
   - images with photometric zero-points derived from PSF-fit photometry; with bit-mask images
   - two source catalogs per image: PSF-fitting and aperture photometry:
   - difference images with QA metadata

   ➔ **public (TBD)**
2. **Reference images (co-adds), coverage, unc maps, and two source catalogs per image**: PSF-fitting and aperture

   ➔ **public (TBD)**
3. **Match-files per readout-quadrant from source-matching of epochal extractions:**
   - based on epochal PSF-fit photometry catalogs: to support "object-based" lightcurve database:
   - object cone searches via user interface ➔ LC + LC-collapsed metrics extracted from source match-file

   ➔ **public (access through user-interface to LC DB)**
4. **Products to support near real-time discovery:** *thresholded* transient candidates (point sources and streaks) with metadata and image cutouts

5. **Historical (users) database of all transient candidates and metadata generated from real-time pipeline**

6. **To commence following survey start**: alert (event) stream extracted from real-time pipeline with metadata

   ➔ **public (TBD)**
7. **Products to support Solar System/NEO discovery and characterization:**
   - moving object tracks from linking point-source transients; known objects are tagged.
   - delivered to the IAU's Minor Planet Center following human vetting.

ZTF Public Delivery Schedule (provisional)

- **First data release**: survey start + 12 months: ~ Dec 2018
- **Second data release**: survey start + 18 months: ~ Jun 2019
- **Third data release:** survey start + 24 months: ~ Dec 2019
- **Fourth data release:** survey start + 30 months: ~ Jun 2020
- **Fifth data release:** survey start + 36 months: ~ Dec 2020

- **Survey start:** ~ late 2017

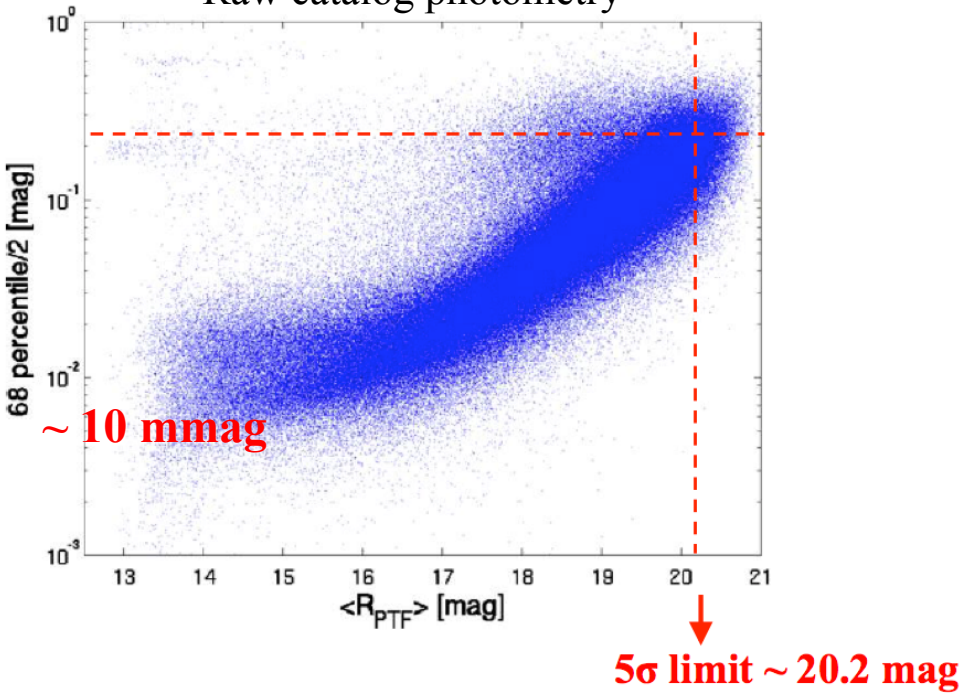
- **Core deliverables for the above:**
  - epochal science images + catalogs + ancillary products (metadata)
  - co-add images + catalogs + ancillary products (metadata)
  - above products searchable through user-inteface according to spatial constraints/survey parameters
  - lightcurve access through user-interface (refined source-matching from epochal PSF-fit catalogs)

ZTF Lightcurve Pipeline

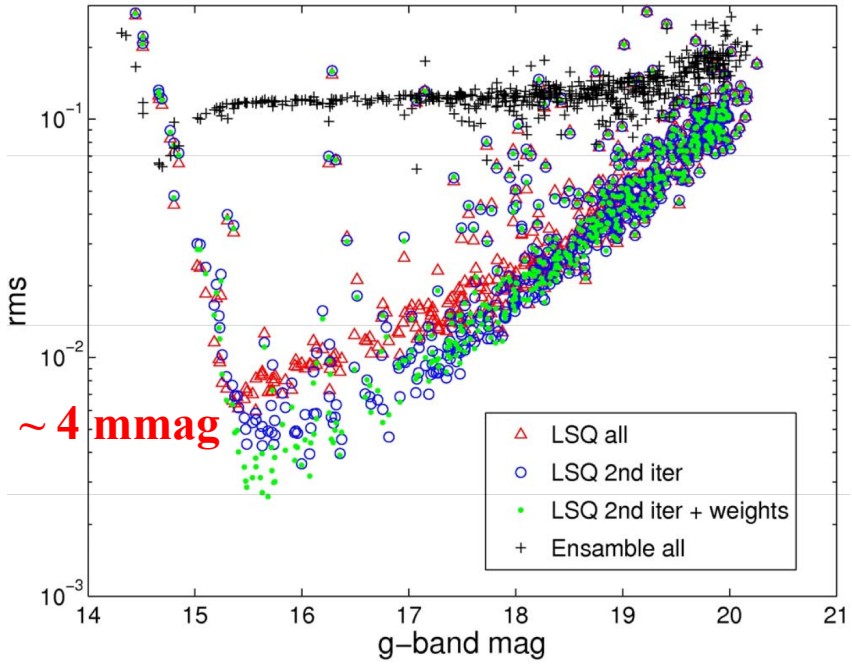
- All sources detected in epochal images are matched against the reference-image source catalog for a given field, CCD image quadrant, and filter
- The "cleanest" least variable sources are used as anchors for the relative photometric calibration
- Individual image gain-correction factors are computed using a least-squares fitting method
- These gain-correction factors are applied the image photometric zero-points
- The refined zero-points improve relative photometry to a few millimag for bright sources
- This pipeline will be triggered on timescales of typically 2 to 3 weeks (TBD)
- All lightcurves for a single CCD image quadrant and filter are stored in a "matchfile" (hdf5)
- Accompanying each lightcurve is a set of >100 metrics: RMSs, Skews, Stetson indices …
- All lightcurves and metrics are seeded by an object ID; these IDs with positions are loaded into a database to support spatial searches; associated lightcurve is retrieved from the "matchfile"
- Expect of order 1.3 billion objects (individual lightcurves) for ZTF

PTF Photometric Performance (internal)

Raw catalog photometry

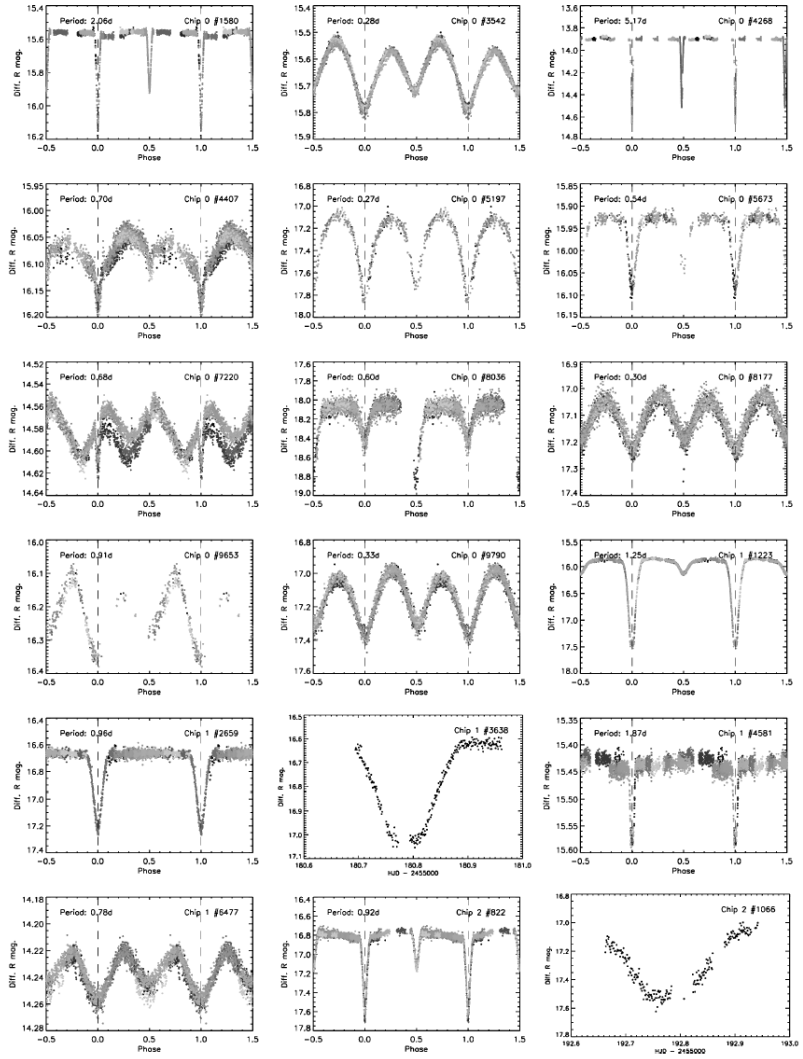


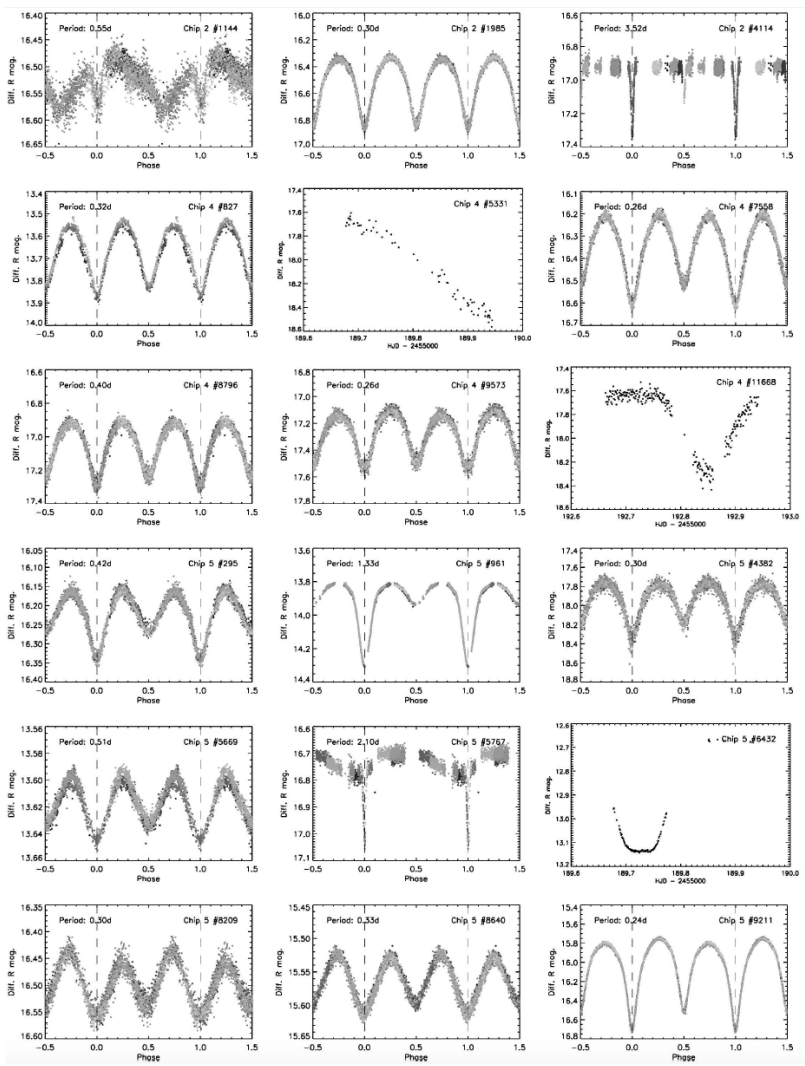
Refined (relative) photometry for lightcurve DB



Example PTF lightcurves from the Orion Project

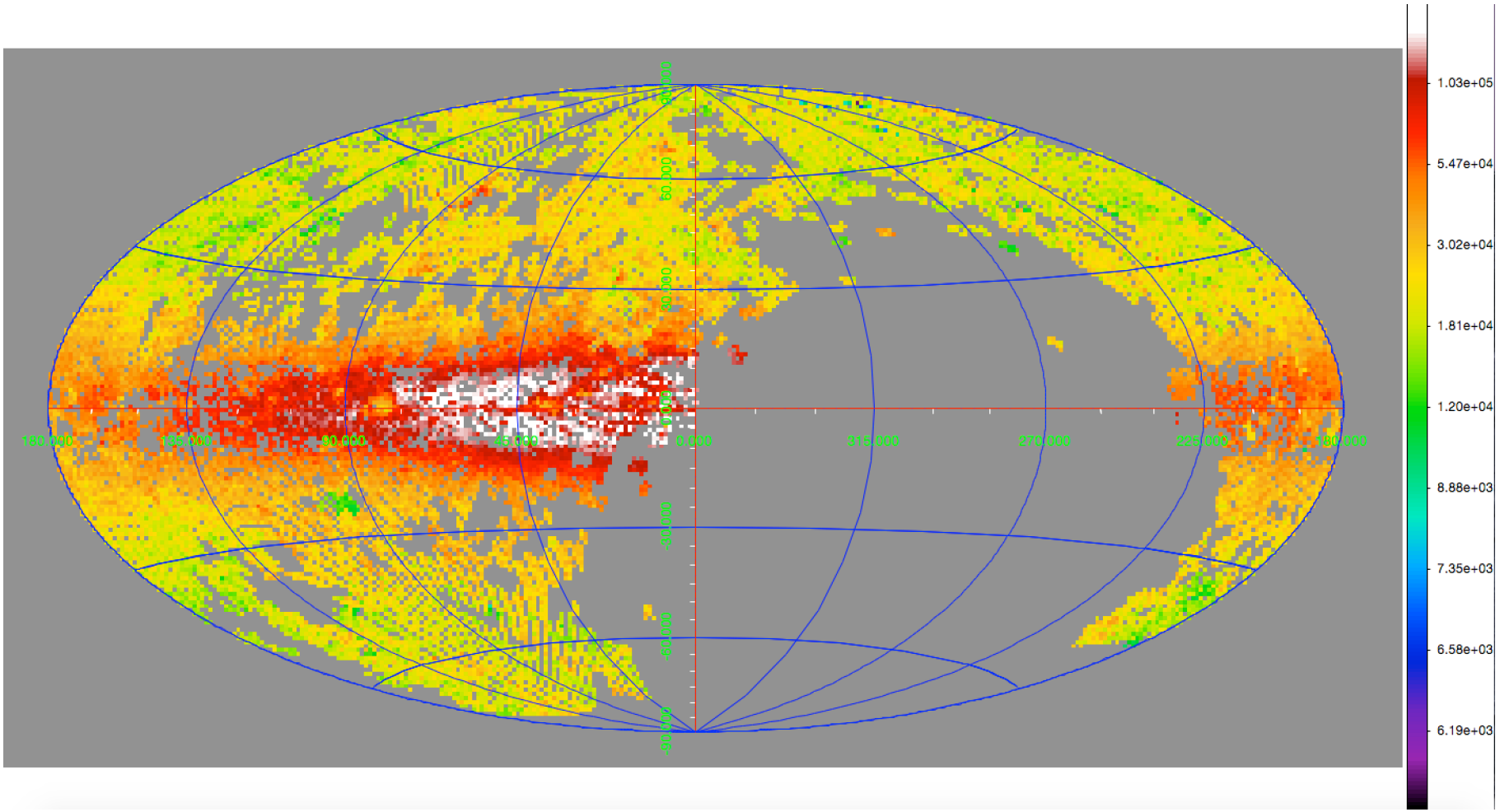
Binary star lightcurves from PTF (Van Eyken et al. 2011)





Density of PSF-fit extractions from PTF CCDs

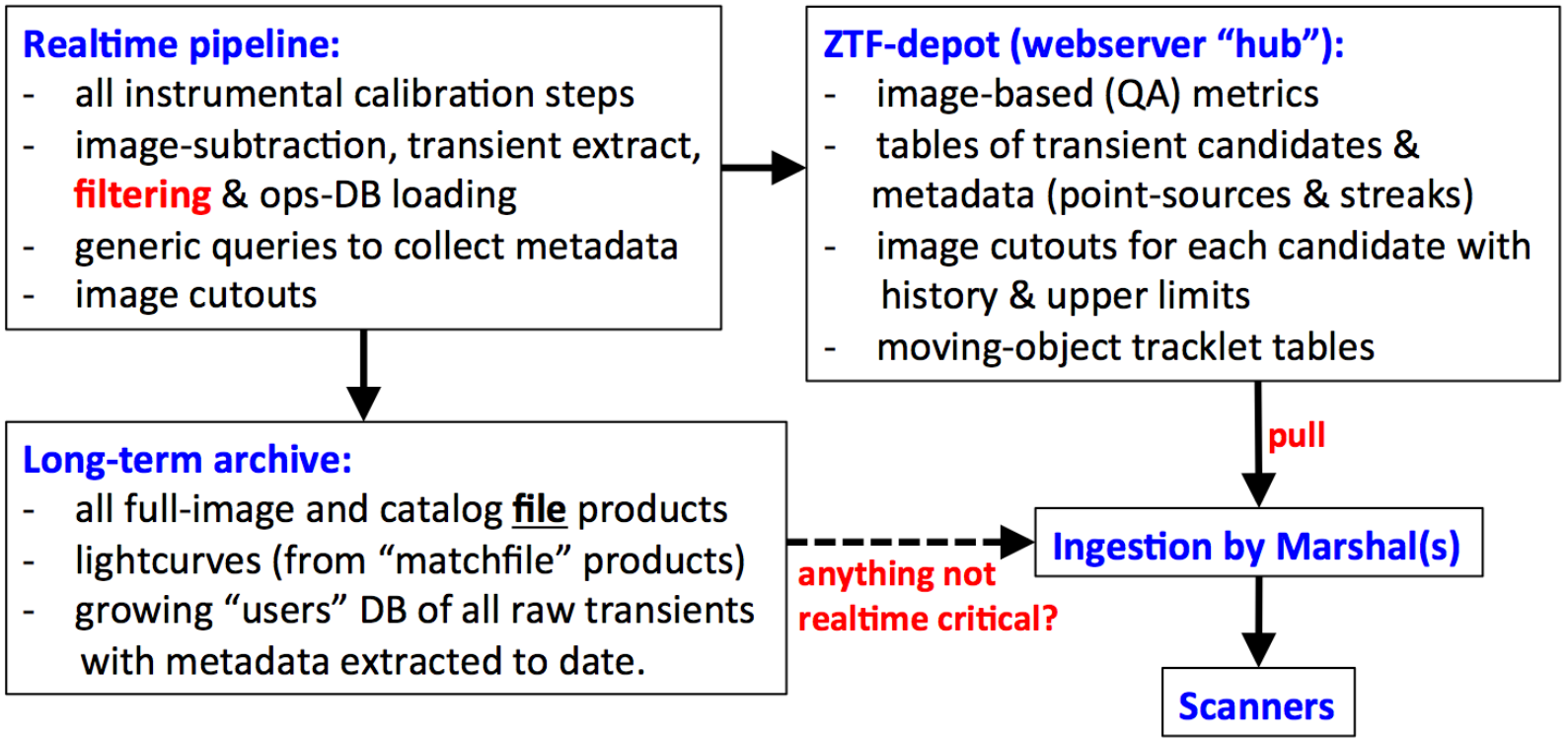
Extrapolating from PTF, we expect of order 900 million PSF-fit extractions **per night** for ZTF(!)



Back up slides

Archive versus Depot products

- To support "fast response science": plan is to deliver a generic event stream (following any automated real/bogus filtering in pipeline) to a webserver for collection by all marshals.
- Other (historical) products, including all extracted events can be retrieved from growing archive.
- Public access to historical database that stores all extracted events/transients is TBD.



ZTF real-time pipeline

- Does most of the heavy-lifting in real-time.
- Time-critical: to support near real-time discovery; fast response/follow-up science
- **Requirement:** 95% of the images received at IPAC must be processed with transient candidates presented to science marshals in <~ 10 minutes (goal is 5 minutes)

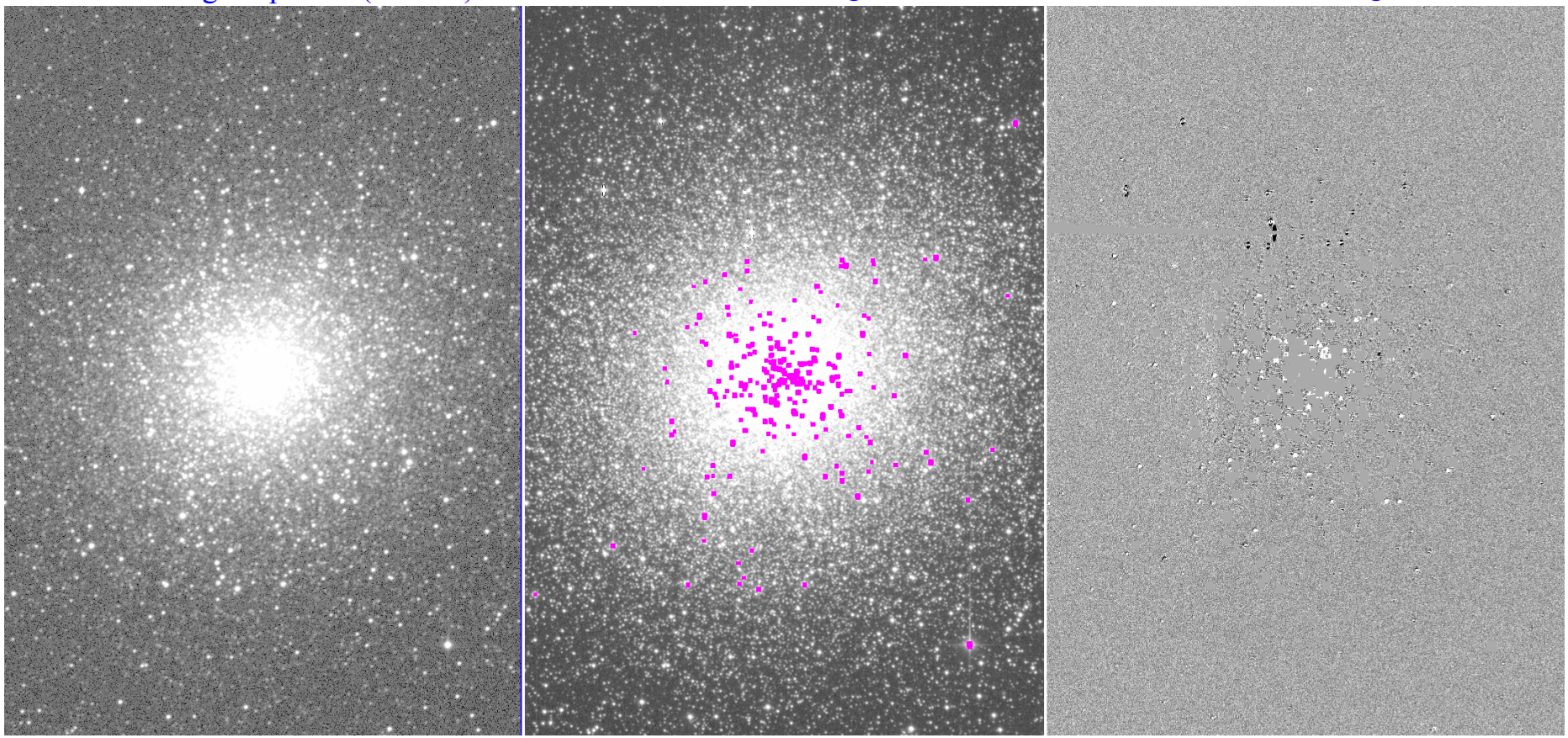
- Real-time pipeline consists of two phases:
  1. Instrumental calibration (bias-corrections, flat-fielding, astrometry, photometric calibration, pixel masks …): generates single-epoch image and catalog products for archive.
  2. Uses outputs from 1 to perform image subtraction, extraction of transient candidates, metadata, cutouts …

- Currently being tested using a camera-image simulator:
  - Takes as input a "schedule" of camera pointings from the survey simulator, with multiple epochs on same region of sky, in any filter.
  - Sources from PanSTARRS1 (DR1) catalog are injected with ZTF instrumental noise.
  - Point-source and streaking transients are also simulated.
  - Raw data files are packaged and compressed according to camera-software specifications.

M13 globular cluster from PTF

- **Enormous benefit:** image-differencing suppresses regions with high-source confusion
- Improves ability to discover flux variables and transients

Science image exposure (R-filter) – Reference image = Difference image



- Bad / saturated pixel regions: colored magenta (zeroed in difference)

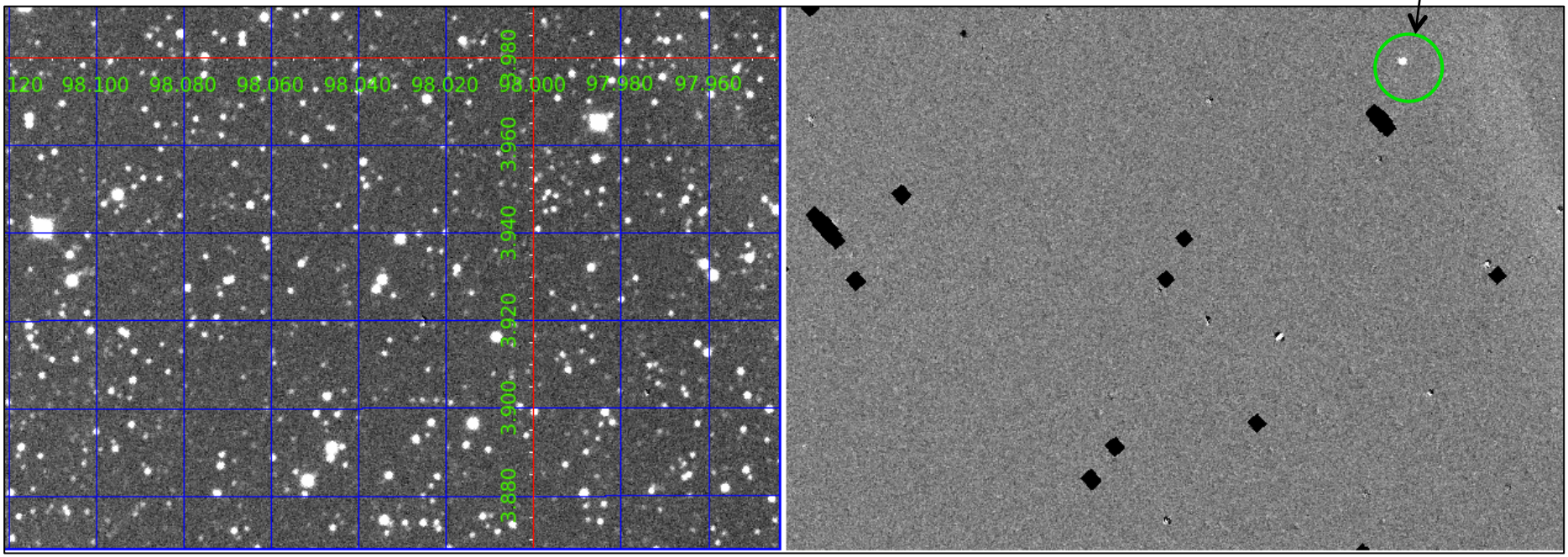
“Good” difference in Galactic Plane from PTF

When upstream astrometric/distortion calibration is near perfect, it works!

known variable

science image exposure (~ 10’ x 7’ *zoom*)

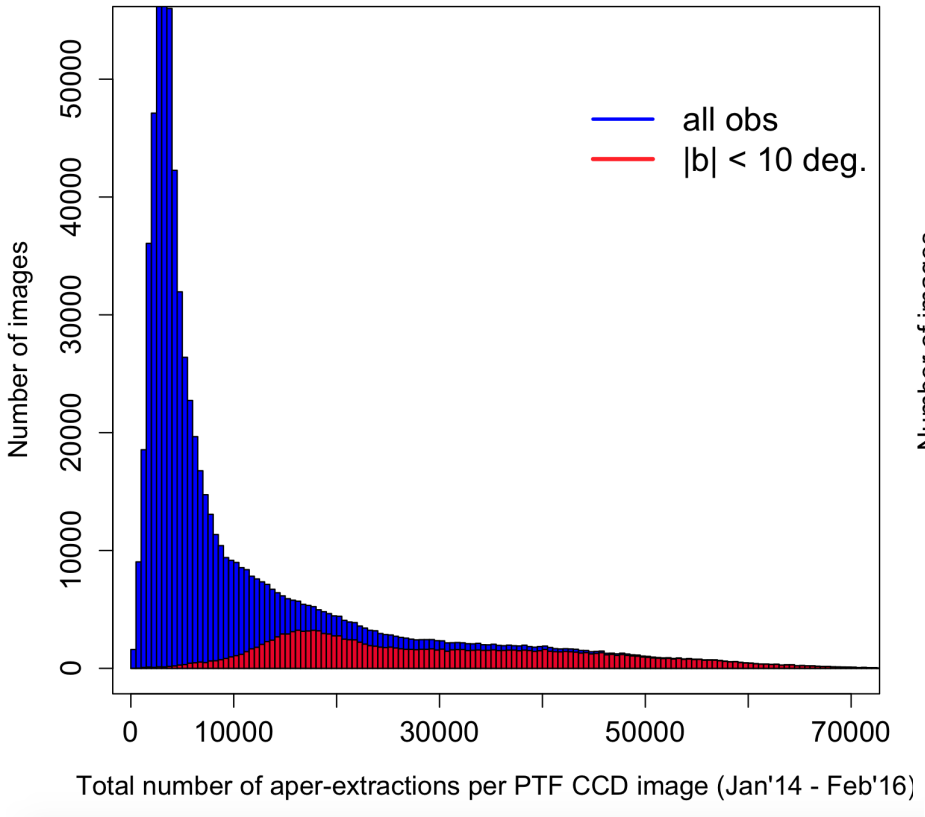
Sci – K (x) *Ref difference image*



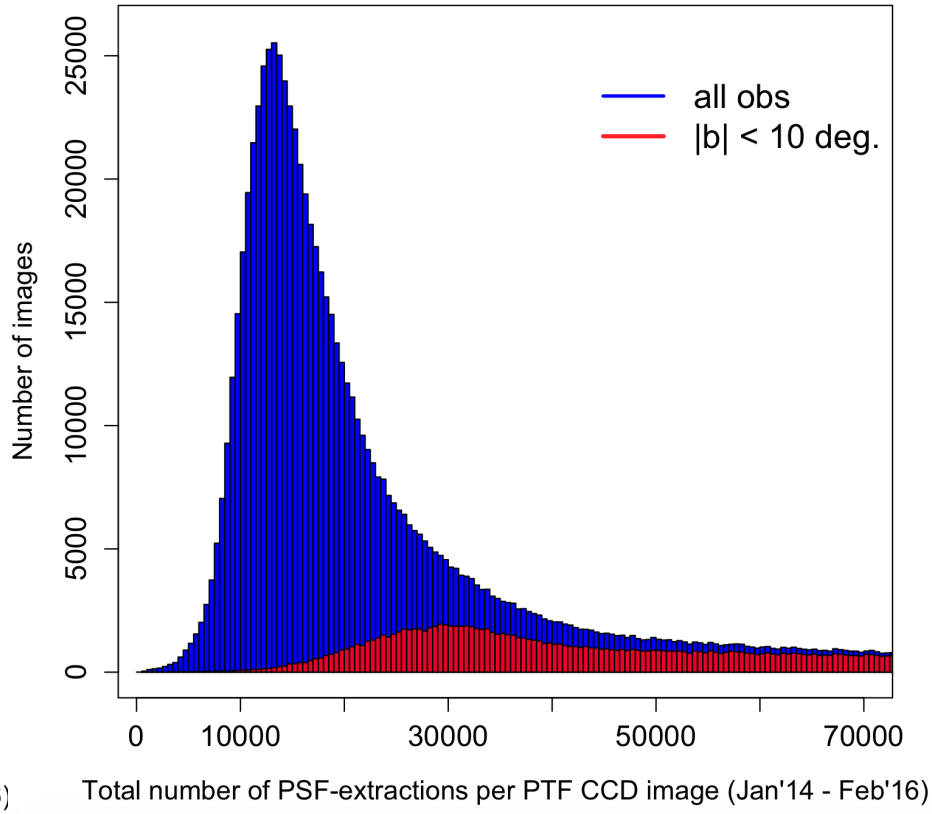
coordinate grid is galactic

Number of sources extracted from PTF CCDs

Aperture (SExtractor)



PSF-fitting (DAOPhot)



ZTF data product volumes / source counts

Per night:

Assuming average length of night at Palomar is ~ 8h:40m (summer: ~6h:20m, winter: ~ 11h), we expect ~ 700 camera exposures per night on average => 44,800 readout quadrant images.

- raw data (including calibrations): ~ 367 GB compressed (3x)
- instrumentally-calibrated epochal images, masks, and metadata: ~ 3.1 TB
- aperture photometry (epochal) catalogs: ~ 140 GB
  - ~ 310 million sources per night
- PSF-fit photometry (epochal) catalogs: ~ 44.8 GB
  - ~ 900 million sources per night
- image-subtractions and metadata ~ 1 TB

Total per night: ~ 5.65 TB

For three-year survey:

Assuming ~ 250 to 280 "good" nights per year (from PTF),

Total image/catalog file products: ~ 3 PB

ZTF Galactic Plane Survey

Galactic Plane Survey

The footprint for the Galactic Plane Survey is the strip of fields within +/- 6 degrees of b=0 (and North of ~Dec -25 or so, TBD). Whenever a field in this footprint is up (e.g., above airmass 2 for > 1 hour), observe it every night, once in g and once in r separated by at least 40 minutes.

The amount of Plane area satisfying this critereon ranges seasonally from about 1000 to about 2000 square degrees.

Source: Project P.I. (at Jan 2017 AAS meeting)